

Facing up to the Chernobyl accident

It is natural that the worst nuclear accident ever should have cast a long shadow, but last week's commemoration of the event and its likely repetition may have gone too far.

If the Soviet Union was not suffering such turmoil and if *glasnost* had not been dimmed by the events of the past few months, it would by now have been possible to piece together an account of what really happened at Chernobyl and in its aftermath five years ago (see page 4). But even now, there is no reason seriously to doubt the account given by the Soviet authorities six months after the accident, at the meeting organized in Vienna by the International Atomic Energy Agency. The cause of the accident was an ill-planned and unauthorized experiment carried through incompetently in circumstances in which malfunctioning nuclear reactors would not have been considered more remarkable than, say, trucks that had broken down.

The same slackness accounted for the failure to foresee the seriousness of the accident, or even to over-react to it, with the consequence that large numbers of people living nearby were needlessly exposed to radiation. The steps taken permanently to isolate the reactor were heroic, but in retrospect may have been characteristically hurried. The exposure to radiation of people living far from the site in the Ukraine, and especially in Byelorussia, has not been disclosed with the clarity its potential seriousness demands. The technical deficiencies are those with which the Soviet Union is struggling, still ineffectually, in all industrial fields. The others are consequences of the Soviet difficulty, five years ago, of facing unpalatable truths.

Matters have since become worse in many ways, not better. Chernobyl has become an issue on which nationalist interests may seek to best the centre at Moscow, whence last week's mock trials and protest meetings, while the whole affair has become a symbol to people throughout the Soviet Union of the past mismanagement of their affairs. For those of us who live elsewhere, it might seem that the moral is that one should never buy a nuclear reactor from a Soviet organization, or rely on Soviet management for environmental protection, but there is more than that to say. First, the rest of the world needs to know exactly what has happened after Chernobyl so as the more confidently to face the decisions it will soon be making on alternative sources of power. Second, the Soviet Union's own reputation requires a proof that it has dealt humanely with all those affected by the accident.

The remedy is in *glasnost*, which means the technical publication of the details. What has happened to those evacuated from the surroundings of the site, what does cytology say about their exposure in the aftermath of the accident, where are they now and what arrangements have been made for

their supervision and care? Tragic though the circumstances may be, information such as this bearing on the influence of small radiation doses on people's health will become with the passage of time an invaluable resource. The same is true of the records of those more heavily exposed to radiation at Chernobyl. The only sufficient memorial to those who died there is that it should be made public in exhaustive detail.

The consequences of Chernobyl outside the Soviet Union have been what the authorities first claimed — minimal. Contamination of reindeer in Northern Scandinavia has abated. A few spring lambs are still considered to be over-contaminated with caesium at two sites in the British Isles, but only because of the precipitate tenfold cut in European safety levels in the first few days after the accident, since when there have been five further years of mostly trouble-free operation at more than 200 reactors on the scale of that destroyed five years ago. The implications of that experience also need serious consideration. Nobody will again pretend that nuclear power is free from risk, but the past five years have also emphasized what has always been plain — that good management can reduce both the risks of accidents and their seriousness.

Different governments and their public utilities will rightly respond differently to these developments. Some who have always been sceptical of commercial nuclear power will remain so, at least for the time being. Some who were once sanguine may have become sceptical because of Chernobyl. But all of them, and also those wishing to push ahead with the development of nuclear power stations, will be helped enormously by more and better information. After five years, that is the best way of marking Chernobyl. □

Recovery by decree

Is President George Bush more alarmed by the threat of a recession than he has been telling?

THE seven rich countries' finance ministers, in Washington last week-end for a meeting of the International Monetary Fund, were probably startled to be summoned by President George Bush to the White House and asked to reduce current interest rates in concert. It is as if the organizers of the annual meeting of the Federation of American Societies of Experimental Biology in Atlanta last week had been faced with a plea from City Hall also to use the occasion to halt the



growth of the world's population, arguably the most serious problem with roots in biology. In the event, the finance ministers appear to have heard out the president politely; they even issued a mildly approving statement. But there will be no concerted move in that direction.

Nor can there be. As Tolstoi might have put it, during economic growth, industrialized economies are much alike, but in recession they go separate ways, looking out for their own interests. On this occasion, for example, both Germany and Japan are more worried by the damage that would be done by inflation than by recession (the growing unemployment in eastern Germany notwithstanding), and are well within their rights to be so. But even if the finance ministers had done what the president had asked, the threat of a further deepening of the recession would not have vanished.

The theory is impeccable: if interest rates fall, people are more likely to borrow money to found new enterprises, so increasing employment and the output of goods and services. That is what happens when things are going well, so that interest rates are valuable economic regulators at times of growth. But when economies are shrinking, people are not tempted to invest in new enterprises by marginal reductions of the real interest rate (nominal interest rate minus the inflation rate), which now vary from 3 to 6 per cent a year in different countries; instead, they are more concerned that an investment that is unwise, perhaps because the market for what it will produce has disappeared, is a recipe for losing everything. The same people are still smarting from the ample illustrations during the recent unwinding of the 1980s boom of how apparently solid assets — investments in commercial banks, for example — can vanish into thin air.

So how, and when, will President Bush's wish be granted? The usual statement — by a return of "confidence" — is tautologous. The real question is that of what conditions must be satisfied before confidence returns. The passage of time, or forgetfulness, will help, but who can wait that long? In contrast with the early 1930s, the US administration has this time set its face against the protection of its domestic market; exporting to marginally more buoyant economies is a useful safety valve, but not a solution of every recessionary economy's problems. President Bush might more usefully have asked at the week-end whether a further deliberate liberalization of world trade would accomplish the trick he hopes for. That, and the encouragement of technical innovation, are the only remedies in sight. □

Shakespeare's school

There is yet another chance to reform British education, in which Shakespeare is only one component.

PRINCE Charles, the British Queen's heir, who delivered this year's Shakespeare Birthday lecture last week, regrets that the Bank of England has decided that Shakespeare's portrait will not always appear on the face of its £20 banknotes. The regret, no doubt, would have been well received by an audience of shakespeareans, but Prince Charles did not tell it that

Shakespeare is to be replaced by Michael Faraday, another English genius. Would that have made a difference?

Most probably, not. The first part of the prince's address was a splendid evocation of Shakespeare, the second a comment on the state of British public education which has been widely regarded as politically controversial (but which is hardly so). The two themes were linked by the argument that Shakespeare's unique insight into the human condition was, by good fortune, written in English, for which reason English-speaking people and the British in particular have a unique component of their cultural heritage to cherish, to cultivate — and to teach in the schools. There is nothing controversial in that, or in the prince's assertion that culture should be a crucial part of general education. So much must be generally agreed.

What is most wrong with British school education outside Scotland (which has more liberal arrangements) is that it is not general education at all, but for many students an inadequate preparation for a life of specialism. It is commonplace that young people with an interest in science are required to commit themselves for, or perhaps against, when they are younger than sixteen; changing course later is difficult, often impossible. But the system is also hag-ridden by examinations, among which the dominant are the examinations (called A-levels) in which success is required for entry to universities and higher education. Only in the past few years has this system been broadened by the introduction (in this academic year) of less demanding courses that universities and polytechnics will accept. But even now, many students will be encouraged by their schools to stick with the traditional pattern, while the educational system is only now waking up to the need that courses in higher education should last for four years, not three. (The latest voice in support of that was a report from the Advisory Council on Science and Technology published earlier this week.) Can anybody wonder that the recruitment of young people into science and technology remains a British headache?

Both Prince Charles's proper interest and that of science and technology would be met if some British government (why not the present government?) were radically to transform and rationalize the system by requiring that there should be no one-to-one link between schools and higher education. Then schools could provide general educations and examinations that were proofs of competence. Universities and polytechnics might have to devise other ways of fitting would-be students to the courses they offer, or switch to open access as others do. But there are two obstacles. One is the peculiarly British conceit that it is feasible to measure attainment so accurately that a young person's educational destiny can be accurately foretold (which rather defeats the objectives of education).

The other, clear from Prince Charles's speech last week, is the contrast between culture and other kinds of intellectual work, in which the former is preferred. Although the Bank of England's decision to replace Shakespeare by Faraday has been compelled the need to redesign all banknotes to be the same size, the switch may help dispel the entrenched conceit. □

US state universities face massive cuts

- Recession brings layoffs, programme closures
- Long-term effect on research still unclear

Washington

A RASH of state budget crises is undermining a decade of US academic growth, as more than half of the university systems in the 50 states now face serious cutbacks and deficits.

Over the past few months, 26 state university systems have been forced to cut their budgets for the 1991 fiscal year — money that had already been committed to salaries and current programmes. Next year, matters look even worse. At least eight state governments are planning to cut their university systems below this year's level — as much as 20 per cent below, in the case of the University of Massachusetts system. Another 16 states do not expect to keep up with inflation in their higher education funding, even in the face of larger numbers of students entering the university systems.

Two years ago, state universities were riding on the crest of a wave of increased enrolment, new programmes and construction that came with the academic boom years of the 1980s, a period in which the state university systems grew on average by 35 per cent.

In 1989, more than three-quarters of all US higher education students attended state colleges and universities.

But then the recession struck, leaving many universities back at 1980s-level funding despite increased enrolments and classes. Most systems have responded by cutting administrative and support staff, postponing construction and equipment purchases, raising tuition and student fees, delaying salary increases, offering early retirement to tenured faculty and eliminating some departments and courses.

It is hard to know what effect the current crisis will have on long-term academic quality. Most university research programmes are supported on federal grants and are largely sheltered from the fiscal storm. Of the programmes that are being cut, many are in fringe or shrinking fields, such as home economics or traditional agriculture. In times of plenty, universities protect such programmes to avoid faculty and student protest. But in lean times, when the need is clear, it is easier to make tough decisions without triggering student demonstrations.

If restructuring trims university programmes back to a lean core of key disciplines, that may not necessarily be a bad thing, despite the loss of academic diversity. "We can't be all things to all people anymore," says Donald Langenberg, chancellor of the University of Maryland. "We've got to focus on what we can do best. Now we can make the cuts we should have made all along."

Like universities elsewhere across the country, University of Maryland officials are planning to eliminate faculty and staff positions, close programmes and departments, and increase class sizes. A 'strategic redeployment' is expected to eliminate two colleges — Human Ecology (what used to be known as Home Economics) and Library & Information Service — and eight other programmes, including nuclear engineering and hearing and speech sciences.



University associations blame the trouble on the nation's general economic malaise, coupled with the Reagan Administration's eight-year policy of transferring responsibility for federal social programmes to the states. Just one of those transferred programmes — Medicaid, health-care support for the poor — will cost the states some \$67,000 million in 1995, more than twice what they spent last year. Facing political pressure against increased taxes, the states were forced to dip into general funds, out of which the university systems are funded.

"Higher education is a big piece of the pie and it's discretionary. That makes it easy to cut," says University of Maryland spokesman John Lippencott. Maryland's higher education budget consumes 10 per cent of the total state budget — a tempting target, and one that did not go unnoticed. Last month, Maryland state legislators cut \$67 million — 12 per cent — from the state university budget.

Cuts at other state university systems include:

■ **California:** In the California State University system, which faces a \$402 million cut from its 1992 request, officials are planning a 20 per cent student fee increase and layoffs of 864 staff, instructors and part-time faculty. They intend to lose 420 faculty positions by offering early retirement. Chancellor Ellis McCune told the state assembly that

in his 30 years within the CSU system, "I have never seen a budget that promises to be as devastating."

Legislators have cut \$295 million from the 1992 request of the University of California system. In response, UC officials are planning to cut 1,000 staff and defer faculty salary increases for six months. They expect to reduce enrolment by 5,500 students, resulting in the elimination of 360 faculty, 100 teaching assistants and 300 support staff. Officials also plan to cut \$20 million worth of programmes that are yet to be named. A one-time 'Golden Handshake' early retirement offer earlier this year was accepted by 3,000 staff, about half of those eligible.

■ **Massachusetts:** The University of Massachusetts system faces a \$115 million cut from its 1992 request. Governor William Weld plans to close as many as five colleges, eliminate 1,000 faculty positions and increase tuition by one-third. The university plans to kill its Community Health Education Program and part of its School of Education. Spokeswoman Kay Scanlan says, "Before the legislature started on the new budget, we had asked all our departments to turn in budgets reflecting a 7 per cent decrease from last year. Now we don't even know what to ask for. Forget keeping up with inflation."

■ **New York:** The State University of New York (SUNY) system request for 1992 has been cut by \$60 million, forcing a 30 per cent increase in tuition fees and 2,300 personnel cuts. At the City University of New York (CUNY) system, a \$92 million cut has resulted in doubled student fees and class closures at six of 21 campuses. CUNY may lose 810 faculty and staff positions at a time when enrolment is up 15 per cent. W. Ann Reynolds, chancellor of CUNY described the predicament as 'the toughest fiscal situation since the late 1970s.'

■ **New Jersey:** State universities have weathered \$72.5 million in cuts over the past three years. Rutgers University plans to cut programmes, 200 staff and 160 teaching positions to make up its largest ever \$38 million reduction this year. The New Jersey Agricultural Experiment Station, located at Rutgers, will lose \$6.6 million.

■ **Oregon:** A state-wide property tax cut has meant an \$86 million loss to the university system in an otherwise healthy Oregon economic climate. The reduction will mean the elimination of 700 faculty and staff positions and some \$50 million in programme cuts.

Portland State's Health and Human Performance School will be closed and its Centre for Public Health Studies is to be absorbed by another department. The University of Oregon will close its College of Human Development and Performance, and mathematics, English, science and speech teacher education programmes.

■ **Virginia:** The University of Virginia system, which faces a \$46 million cut, intends to let attrition eliminate 153 faculty jobs, and to impose a 2 per cent salary cut on all remaining faculty.

Christopher Anderson

World researchers to take a closer look at Chernobyl

Munich

FIVE years after the nuclear disaster at Chernobyl, the world may finally get a closer look at the health effects of the radiation released by that accident.

Until recently, the Soviet Union has not been cooperative about releasing details of the calamity, but now, in conjunction with a programme of the World Health Organization (WHO), it will apparently open its doors to health researchers from the rest of the world.

On 6 May, the World Health Assembly, which is WHO's parliament, is expected to vote to lend WHO support to an international research programme to study the effects of Chernobyl. The programme, which began last year as a Soviet pilot project, will allow international researchers to gain unprecedented access to the Soviet population in order to carry out epidemiological and other studies and to track the effects of the accident on the incidence of cancer and other diseases.

A year ago, the Soviet Union began to set up a research centre at Obninsk, 100 km southwest of Moscow, to coordinate research on the effects of the accident. The Obninsk site was chosen because it already

houses 15 research institutes of the Soviet Academy of Sciences that deal with various aspects of radiation and nuclear science. At the same time, Soviet officials also began to set up satellite research centres nearer to Chernobyl itself in each of the three republics most strongly affected by the accident, the Ukraine, Byelorussia and the Russian Federation.

Earlier this year, Japan became the first WHO member state to volunteer assistance for the programme, pledging \$20 million to improve the facilities at Obninsk. The pledge came in response to a letter to the Japanese government written by Hiroshi Nakajima, the Japanese director-general of WHO, which is based in Geneva.

And now that the World Health Assembly has given its stamp of approval to the programme, WHO officials say that other members of the organization (which has 166 member states) are more likely to follow the Japanese lead and make donations. WHO itself has offered to play a coordinating role for international cooperation in the Chernobyl effects study programme, but it is not in a position to devote any of its own budget to the programme.

Long-term support for the programme is

an absolute necessity, WHO officials emphasize, because many of the cancer cases will not emerge for another 5-15 years or more.

The programme will attempt to track the progress of several types of cancer, the incidence of which is expected to rise because of exposure to radiation. The first cancers and other abnormalities to emerge in the next few years are expected to be leukaemias, followed by thyroid disorders after a 15- to 20-year latency period.

Another focus of the programme will be to assess the psychological and psychosocial effects of the disaster. Inspection teams from WHO and the Vienna-based International Atomic Energy Agency have reported that the largely rural local population blamed a variety of illnesses, especially among children, on radiation, even when this was clearly not the cause. The reported increase in children's illness, and in high blood pressure, anaemia and pulmonary disorders in the adult population are thought to be due more to feelings of depression and helplessness among the population than to the effects of radiation.

One potential stumbling-block for the programme is that it may not be able to determine how many cancers are due to the Chernobyl accident because Western experts say there were very few data available from the region before 1986 to use as a baseline for the study.

Steven Dickman

Five-year toll: 10,000 dead from Chernobyl?

London

How many people have really died because of radiation released by the Chernobyl accident? A Soviet scientist has reopened that question with his claim that the Soviet Union has covered up thousands of deaths, but other scientists remain sceptical.

Vladimir Chernousenko, a member of the clean-up team sent into the Chernobyl area after the accident, claims that between 7,000 and 10,000 of the 600,000 so-called 'liquidators' brought in to clean up after the accident have died because of exposure to radiation. The official death toll is 31. The claim, made by Chernousenko in a British television documentary programme, broadcast by Thames Television, has been dismissed by the Soviet Ministry for Nuclear Power and Industry, and questioned by Western experts who have studied the consequences of the Chernobyl accident.

John Gittus of the British Nuclear Forum, who led the UK Atomic Energy Authority team studying the consequences of Chernobyl, says that non-acute health problems caused by radiation should be impossible to detect only five years after the accident. For solid cancers, studies of radiation-exposed populations, such as the Hiroshima and Naga-

saki bomb survivors, show a ten-year latency period before excess cases appear. The UK Atomic Energy Authority estimates that Chernobyl will eventually cause 10,000 excess cancer deaths in the Soviet Union. This represents only 0.03 per cent of the natural Soviet cancer rate.

Evelyn Sokolowski, from the Swedish Nuclear Training and Safety Centre, says that the death toll given by Chernousenko is consistent with the natural mortality rate in the Soviet Union among a population of 600,000 averaging 29 years of age.

In an apparent attempt to discredit the cancer claims, Soviet authorities initially said that Chernousenko had no scientific credentials to discuss the effects of Chernobyl. In a biography distributed by the documentary's producers, Chernousenko had been described as the scientific director within the 30-km exclusion zone that surrounds Chernobyl, and deputy chairman of the Ukrainian Academy of Sciences Commission responsible for rectifying the consequences of the accident. The Soviet authorities denied that Chernousenko was even a scientist.

Those authorities have now issued a second statement, quoting Victor Bar'jakhtar, chairman of the Ukrainian

Academy Commission on Chernobyl. This acknowledges that Chernousenko is a senior scientist at the Ukrainian Academy Institute of Theoretical Physics and a consultant to the Commission on Chernobyl, but says he is not the commission's vice-chairman.

Despite the general rejection of Chernousenko's fatality claims, some scientists who have visited the Chernobyl site are concerned about the safety of the Soviet scientists still working at Chernobyl — some are even working inside the concrete 'sarcophagus' built to enclose the shattered reactor.

Brian East, head of health physics at the Scottish Universities' Research and Reactor Centre, says that Soviet physicists are working in primitive plastic overalls in an area likely to be rich in plutonium, and seem to have no equipment to measure surface contamination with alpha particles; in the United Kingdom, workers entering areas where plutonium is being manipulated wear sealed suits with a positive pressure.

A more complete analysis of the health and environmental effects of Chernobyl will be released later this month by the International Atomic Energy Agency, which has co-ordinated a study involving a number of United Nations organizations.

Peter Aldhous

Pollution on the upswing

Tokyo

JAPANESE industry has set an example for the world by reducing its energy consumption and carbon dioxide emissions through energy-efficient technology and production processes. But Japan's 120 million citizens and the nation's trucking companies are rapidly undoing all this hard work. According to the latest annual white paper (policy document) on the environment, released last week by the Environment Agency, Japan's consumer and transport sectors are steadily increasing their consumption of energy and emission of pollutants — cancelling out the environmental gains made by Japanese industry.

Although the small and comparatively young Environment Agency has little political clout in the Japanese government, it has become increasingly bold in its pronouncements, reacting to growing concern about the global environment. And its recent white papers have attracted worldwide as well as national attention.

The white paper, in what the agency claims is a first for any country in the world, breaks down carbon dioxide emissions in Japan into four sectors: industry, private households, transport and energy conversion and electric power transmission losses. The breakdown provides a revealing insight into where Japan's carbon dioxide problems lie.

Analysis of emissions over the past 25 years reveals that carbon dioxide discharges from industry fell significantly after the 1973 oil crisis, as industry rallied to improve its energy efficiency. Industry's share of total emissions has accordingly dropped from 60.2 per cent in 1973 to 49.5 per cent in 1989. And this was achieved against a background of a rising gross national product and industrial production (except for a brief period of negative growth immediately after the oil crisis).

On the other hand, consumption of energy and carbon dioxide emissions in private households have continued to rise steadily. Households now account for 23.2 per cent of all emissions, compared with 18 per cent in 1973, and their share of electricity use has also expanded dramatically as more and more Japanese homes have installed air-conditioning units and other electrical appliances.

Similarly, the share of emissions from the transport sector climbed from 13 per cent in 1973 to nearly 20 per cent today. According to the white paper, the biggest culprits are the small- to medium-sized companies that operate the thousands of diesel trucks that now jam Japan's narrow roads. The trucks not only emit carbon dioxide but also contribute much of Japan's nitrogen oxide (NO_x) pollution.

Paradoxically, Japan has some of the strictest regulations in the world for control of

emissions from cars. The regulations were introduced because of severe urban pollution problems in the 1960s and early 1970s, and, as a result, NO_x pollution in Japan's cities decreased steadily from a peak in the mid-1970s until the mid-1980s. But the white paper shows that since 1985, NO_x pollution has been on the rise again, and the major source of this pollution is diesel trucks.

Emission controls for diesel engines are not so strict as for gasoline-driven vehicles and, because diesel fuel is considerably cheaper than gasoline, there has been a boom in the use of diesel trucks among small trucking companies, which have expanded their operations dramatically in recent years. The boom in the trucking business has been caused by increased demand for private parcel delivery services and by the general expansion of consumption of products in Japan's increasingly affluent society.

Ten years ago, the millions of small convenience stores and groceries that line the streets of Japanese cities managed with only one or two deliveries by truck per day. But now, as manufacturers compete ferociously to grab a share of the market, dozens of deliveries in a day are not uncommon. One example, cited in television reports last week, is the current 'beer war', where beer manufacturers have flooded the market with 174 marginally different brands of lager beer — all of which need transporting to retail outlets.

The problem is compounded by Japan's ENVIRONMENT

One Westerner gets it RITE in Japan

Tokyo

LAST week, when the Japanese Research Institute of Innovative Technology for the Earth (RITE) announced the 12 winners of grants for research on the global environment, 11 were Japanese scientists, even though the programme was in theory open to researchers of all nations. The result is not too surprising, however, given the way in which the grants were publicized outside Japan — that is to say, not at all.

The episode illustrates how the Japanese sometimes send conflicting signals to the rest of the world — in this case, seemingly asking for outside participation but discouraging it at the same time.

RITE was established last year by the Ministry of International Trade and Industry (MITI) to develop 'environment-friendly' technology (see *Nature* 350, 266; 1991). As part of its task, it offered the 12 grants, which range in value from ¥3 million to ¥7 million (\$22,000–\$50,000) a year. MITI decided that the institute should open up the programme to non-Japanese as part of a policy to internationalize MITI's research projects.

grossly inadequate road system. Tokyo has a network of toll-charge 'expressways' built in the 1960s. But these major arteries are only two lanes wide and at some key junctions they narrow to one lane. Traffic jams that extend for miles are an everyday event. And the jams are getting worse every year as more trucks and cars flood the roads.

Kei Takeuchi, an economist at Tokyo University's Research Center for Advanced Science and Technology, who leads a team of researchers trying to find social and economic solutions to global environment problems, blames the Japanese government for much of the problem. It is all very well for the Japanese automobile industry to offer leading-edge emission controls and fuel-efficient cars, he says, but until the government takes the initiative to solve Tokyo's traffic jams, no amount of technology will ease the city's growing pollution problems.

A few months ago, the Tokyo metropolitan government tried to improve the situation by asking transport companies to refrain from using their vehicles on Wednesdays. But only 5 per cent of 5,500 companies complied.

Toshiro Kojima, director of the planning and research division of the Environment Agency, says the agency hopes that the price differential between diesel and gasoline fuel will be eliminated and tax incentives will be introduced to encourage transport companies to stop using diesel trucks. But because the trucks typically cost about ¥10 million (\$70,000) each and last for years, he admits that a rapid switchover is unlikely.

David Swinbanks

But unfortunately for non-Japanese scientists, the New Energy and Industrial Technology Development Organization (NEDO), in conjunction with RITE, is responsible for funding and publicizing the grants. NEDO is a semi-government organization that runs most of MITI's projects, including RITE, and it does not seem to share the ministry's enthusiasm for internationalizing the projects.

For the RITE programme, English literature describing the grants was sent to foreign embassies in Tokyo, but the only other announcements appeared in Japanese publications and were written, of course, in Japanese.

Very few non-Japanese scientists noticed the announcements for the new programme before the deadline in late January, and only one will receive an award. S. C. Jarvis of the UK Agricultural and Food Research Council's Institute of Grassland and Environmental Research will get ¥6 million to study methods to enhance the capacity of agricultural soils to provide sinks for methane.

David Swinbanks

A global experiment in technology transfer

Irvine, California

IN 1987, after a year of delicate and sometimes acrimonious debate, 31 nations signed the Montreal Protocol, agreeing to phase out their use of chlorofluorocarbons (CFCs) in order to save the Earth's ozone layer. Last summer, after more tricky negotiations, the developed nations even offered to pay part of the cost for the less developed countries. That, however, was just the easy part.

Now, countries around the world are struggling to determine how they will keep their promise to replace the valuable and ubiquitous CFCs. Researchers and engineers are trying to develop refrigeration systems and insulating foams that do not use CFCs. Politicians are debating the trade-offs among cost, convenience, energy efficiency and environmental damage that will have to be made with the substitutes. But perhaps the biggest challenge will be transferring the technology for CFC replacements from developed countries, where it is being devised, to developing countries.

Last week, the US National Academy of Engineering sponsored an international conference to review progress in developing CFC replacement technology and to map out strategies to transfer that technology to developing countries. As was clear by listening to the delegates from two dozen nations around the world, technology transfer is still more an ideal than a reality.

Yet the success of this technology transfer will determine more than just the final score in the battle against ozone depletion. In many ways, the fight for the ozone layer is a trial run for solving more difficult global environmental crises that may lie in the future. As conference participant Anthony Vogelsberg, environmental manager of Du Pont's fluorochemicals division, put it, "If you can't deal with this effectively, then how are you going to deal with global warming or other problems that come up?"

CFCs have a wide variety of uses, including refrigerants in refrigerators and freezers as well as in building and automobile air-conditioning systems; 'blowing agents' for making insulating and other types of foam; cleaning solvents used in the manufacture of computer chips; and the propellant in aerosol sprays. The chemicals, which consist of one, two or three carbon atoms with chlorine and fluorine atoms attached, are particularly stable, which allows them to make their way into the stratosphere, where they break down and release chlorine atoms, which destroy ozone.

Even before the Montreal Protocol, several of the world's major chemical companies were working to find chemicals that can replace CFCs, and it now appears that the replacement of CFCs will go smoothly in the developed world. It is the developing countries that will be the problem. Since the manufacture of CFCs is relatively 'low tech',

many developing countries have their own manufacturing plants, mostly to supply refrigeration needs. They will have to give these up and — at first, anyway — be dependent on the large multinational companies which have the expertise and the capital to set up the plants for CFC substitutes.

The changeover will not be as simple as having the developing countries give up their CFC production and move to the replacement chemicals. Almost all of the equipment that uses CFCs — refrigerators and air conditioners, foam-making machinery, cleaning equipment in semiconductor-manufacturing plants and so on — will have to be replaced or retrofitted. Somehow, engineers and technicians around the world will have to be instructed in the design, manufacture and operation of a whole new generation of equipment.

This will demand technology transfer on an unprecedented scale. The participants at the conference reviewed a number of obstacles that they have encountered or expect to encounter.

■ The companies that are developing the CFC replacement technology wish to protect their patents and intellectual property rights — and make a profit — but developing countries want the technology as cheaply and with as few strings attached as possible. Originally, countries such as India had asked that the new technology be given to them free; however, the technology belongs not to governments but to corporations, which do not give away their assets. Indeed, companies such as Du Pont would prefer not even to sell the technology itself. Du Pont's Vogelsberg says the developing countries should "either buy from world-class plants or, if they want local manufacture, should have joint ventures."

The spectre of the Bhopal disaster at a Union Carbide plant in India is never far from anyone's mind in such discussions, and Vogelsberg says Du Pont would be unwilling to operate a high-technology CFC replacement plant in a less developed country unless the company had a controlling interest in order to guarantee the safety of the plants.

In countries such as India, where the government restricts foreign ownership, the multinationals are likely grudgingly to licence their technology, but that brings up another problem. The developing countries will ask the multilateral fund to pay their licensing fees, but \$160 million will go only so far. At some point, the corporations can expect to be pressed to lower their fees on CRC replacement technology.

Furthermore, the corporations will be exceptionally cautious about selling their latest and best technology to countries such as India and China, both of which have reputations for not being completely respectful of intellectual property rights.

■ Many developing countries do not have

the capital to finance the changeover to a totally new technology. A multilateral fund established last year by the signatories of the Montreal Protocol will pay only the extra costs associated with replacement technology, not the entire cost of new facilities. If, for example, a developing country wishes to build a refrigerator plant to take the place of an existing facility that makes CFC-based refrigerators, it must pay what a new CFC facility would have cost; the fund makes up the difference. But the developing nation must still invest a large amount of capital in the new plant — and swallow the costs of shutting down the old one.

One solution is for developing countries to invite multinational corporations to build new plants inside their borders in joint ventures, but that can run into the same Catch-22 as above, with both parties wanting control of the joint venture.

■ In countries such as Brazil, which has a history of successful joint ventures with multinational corporations, the problems centre instead on the lack of a government framework for technology transfer. Brazilian industry has already started to plan for a CFC-free future by forming a formal study group to prepare for the change over to new technology, said Alberto Carrizo, director of the engineering, research and development and quality division of Climax Industries in Brazil. The companies are interested in export business and recognize that they cannot afford to fall behind their competitors in the developed countries, he said.

But the Brazilian government has not moved as fast. "Nothing is happening in the government," said Suely Carvalho of the department of experimental physics at the University of San Paulo.

"We can have access to an international fund of over \$100 million, but we do not have the institutional channels." Brazil needs to establish procedures for industries to get access to the multilateral fund, Carvalho said, as well as to put together a central clearing-house of information about CFC replacements.

Eventually, Brazil will want to develop an indigenous capability in this new industry, she said, and for that, "we need an integrated effort among society, research centers, industry and government."

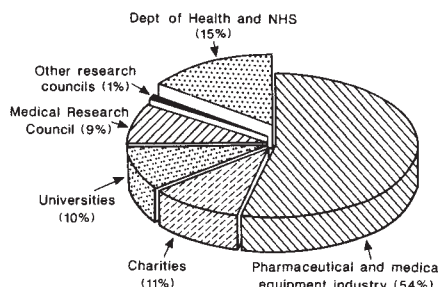
■ The physical task of training workers and technicians in developing countries in the new technology seems overwhelming. Even in countries that import the replacement CFCs, the indigenous industries will need to learn to make new equipment that works on the new chemicals. Refrigerators and air conditioners will have to be redesigned, for instance. And because the technological infrastructure of developing countries is less sophisticated than that of the United States, Europe or Japan, the developing nations generally will have to redesign the technol-

HEALTH RESEARCH

NHS launches first strategy

London

MORE than 40 years after its creation, the UK National Health Service (NHS) last week announced its first-ever comprehensive research strategy. Michael Peckham, director of research and development for the NHS since January, says the aim is to "pervade the NHS with research activity". For the first time, each Regional Health Authority (RHA) will have to draw up research plans



Health research funding in England 1989/90
Source: UK Department of Health.

and show the NHS management that these are being carried out.

Most NHS-funded research projects now are aimed at assessing the effectiveness of different treatment methods. This will be stepped up, says Peckham, to weed out ineffective methods. The NHS will also keep a "weather eye" on fundamental biomedical research, looking for advances that may yield new treatments, he says. But sceptics point out that new and more effective treatments are often more expensive, which may dissuade hospitals from introducing them. Under NHS reforms introduced last month, health authorities can choose to send patients to the hospitals that seem to offer the best value for money. As the health authorities have limited budgets, there are fears that there will be pressure to send patients to

hospitals where treatment is cheapest.

A new central NHS research and development committee is to be appointed in the coming months, and will set between six and ten 'priority areas' for NHS research. RHAs will then bid for grants from a central pot of money to support these research areas. This centrally directed fund should ensure that the separate research plans put together by the RHAs fit within the national strategy, says Peckham.

The first job, however, is to find out just how much the NHS now spends on research — a figure now known only approximately. Peckham estimates it to be about £225 million a year (about 1 per cent of the total NHS budget), but freely admits that NHS research spending has never been calculated accurately. The government plans to increase spending on research to 1.5 per cent of the NHS budget within five years.

Peckham also wants to increase collaboration between the NHS and other British bodies funding medical research. He aims to set up several research liaison groups, to bring together representatives from the NHS, the Medical Research Council, the medical research charities, industry and the universities. Each liaison group will concentrate on a specific health problem, such as cardiovascular disease or mental health. Peckham's model is the UK Coordinating Committee on Cancer Research, run by the Medical Research Council and several of the leading cancer research charities, and widely praised for its role in coordinating British cancer research.

Peckham's appointment to the new post of NHS director for research and development followed criticism from the House of Lords Science and Technology Committee of the low priority given to research in the NHS.

Peter Aldhous

ogy to match their own needs and capabilities.

So far, one success story in the effort to replace CFCs seems to be Mexico. There, according to Jorge Corona, the director of ecological studies for the National Chamber of Industrial Transformation, the government took the lead in deciding to phase out CFCs as quickly as developed countries and industry quickly jumped on board. With the coming free trade pact with the United States, Corona said, "it would be foolish to be 10 years behind our trading partners in developing the replacement technology."

Mexico is phasing out CFC aerosols faster than Japan, and it has cut its use of CFCs in foam manufacture by half, said Geno Nardini of Valvulas de Precision. It is still waiting for adequate substitutes for refrigeration and air conditioning uses, he said, but "local companies are preparing themselves so that when they get the okay from the US and Europe, they're ready to go."

Mexican industry has an advantage over that of many developing countries because of close ties with some of the multinational corporations that will be developing the replacement technology, and it intends to use those contacts for "seed help", Nardini said. But it plans to adapt the technology for its own uses and then help other developing countries to make the changeover from CFCs. "Once technology is adapted to one developing country, it is much easier to transfer to another country at the same stage of development," Corona said.

Like the other delegates at the meeting, Corona and Nardini recognize the importance of establishing the CFC replacement campaign as a prototype for the future. "This has helped the whole world study the mechanisms for fighting this sort of problem," Corona said. "If we had not had the experience we're having now dealing with CFCs, I don't think the world would be ready for dealing with global warming."

Robert Pool

IVORY TRADING

African states invoke origin test to resist ban

Johannesburg

SOUTH African states are planning to market culled ivory which has had its origin certified using a technique described last year in *Nature*. In principle, the technique can ensure that ivory does not come from poached elephants and thus protect Africa's elephant population, while still making some ivory available for sale.

Five states (Botswana, Namibia, Zimbabwe, Malawi and Zambia) are concluding plans to establish a Southern African Ivory Marketing Centre (SACIM), which will auction ivory at a market in Gaborone. They hope to persuade the Convention on International Trade in Endangered Species (CITES) to exempt certified ivory from the international ban on trade in this commodity.

When CITES decided in October 1989 to

declare a two-year moratorium on the ivory trade, it angered Botswana, South Africa and Zimbabwe, which derive considerable income by exporting ivory culled from their populations of African elephants. The populations in these countries, in which poaching is largely under control, have been growing even while the elephant population across Africa as a whole has been in decline — from about 1.34 million individuals in 1979 to 622,000 today. Botswana, Zimbabwe and Malawi entered immediate reservations to the ban, but these are of little avail so long as importers (particularly Japan, which imported almost half the world's ivory in 1989), observe it.

Ivory is currently being stockpiled in South Africa, as culling programmes in the Kruger National Park (which contains most of its elephant population), are con-

tinuing. The National Parks Board will lose an estimated R6 million in revenue over the two-year period. Zimbabwe and Botswana have probably suffered considerably higher losses, and are far more strapped for cash.

Until recently, there had been no way of determining whether ivory originated from a country where culling is permitted or from one where elephants are in danger of extinction. But the authors of two recent papers in *Nature* (346, 744–749; 1990) provided a solution: by measuring the isotope ratios of carbon, nitrogen and strontium in samples of ivory and bone, different isotopic "signatures" can be determined for elephants from different localities.

Botswana's deputy director of Wildlife, Nigel Hunter, says that part of SACIM's revenue will be channelled back into elephant conservation programmes.

Michael Cherry

ESA's eye on the oceans

London

THIS month, the European Space Agency (ESA) should launch ERS-1, the first stage of ESA's Earth observation programme, and the most ambitious package of remote-sensing microwave instruments yet put into orbit. The satellite will provide a wealth of oceanographic data, which will help to improve climate prediction models, but controversy still surrounds ESA's policy on setting prices for those data. Except for some 200 scientists chosen by ESA as "principal investigators" for the mission, academic researchers will be charged for ERS-1 data at the same prices as those fixed for commercial users.

ERS-1, which will fly in a polar orbit, was due for launch on 3 May, from Kourou, French Guyana. But on Monday this week, problems with the third stage of the Ariane-4 launch rocket had set the launch date back, possibly by more than a week.

The key component of the ERS-1 payload is the Active Microwave Instrument (AMI). Because clouds are transparent to microwaves, this will give an uninterrupted view of the Earth below. In its Synthetic Aperture Radar (SAR) mode, the AMI will produce detailed radar images of a 100-km-wide strip of the Earth's surface. These images will be used in land-use surveys, to monitor oil slicks at sea, and may allow oceanographers to detect ocean fronts, where bodies of cold and warmer water meet.

SAR equipment is not new in space. In 1978, the US National Aeronautics and Space Administration (NASA) Seasat satellite included such a radar device. But Seasat

gave only a tantalizing hint of the full possibility of SAR data: the satellite suffered a massive power failure after only 100 days in orbit. The launch of ERS-1 is part of a new surge of interest in SAR imaging. The Soviet Union launched a satellite fitted with SAR earlier this year, and the SAR-equipped Japanese Earth Resources Satellite is due for launch in early 1992.

Unlike the Soviet satellite, ERS-1 will do more than produce SAR images. For most of the time, the AMI will operate in its 'wave' mode, taking snapshot images of a 5-km square every 200 to 300 km along the ocean surface, or in the 'wind' mode, measuring the backscatter from radar beams angled at 45 degrees to the sea surface. This backscatter can be used to calculate wind speed and direction at the sea surface. These data will be a key input to the World Ocean Circulation Experiment, the largest international collaboration in oceanography.

ERS-1 will also carry a radar altimeter, to measure variation in sea level and the height of ice sheets, and the Along Track Scanning Radiometer, which will measure the water content of the atmosphere and sea surface temperature.

ERS-1 has such an impressive range of hardware in part because ESA spending is decided by industry ministers from ESA member states, which are willing to pay into ESA's space programmes because their space technology industries benefit from the manufacturing contracts that result. But this commercial mindset also influences ESA's data policy, so that academics will be charged the same for ERS-1 data (several

hundred pounds for each SAR image) as any other user.

This rigid commercial policy has been pushed most strongly by the British. But some relaxation, perhaps by allowing academics working on environmental research cheaper access to data, is expected for ESA's future remote-sensing satellites. NASA, which has a stronger interest in research than ESA, already takes this line (see *Nature* 346, 600; 1990). The difficulty of selling environmental remote-sensing data is likely to be a major factor in ESA's future data policy.

Many ESA officials say it is not possible to recover costs by selling remote-sensing data — a view which is supported by the disappointingly low number of applications for data received by the commercial US Landsat satellite.

Peter Aldhous

ASTRONOMY

Keck telescope twin gets the go-ahead

Washington

KECK I, which at 10 metres was to be the world's largest telescope when completed next year, has a competitor for the title. The W. M. Keck Foundation announced last week that it has committed \$74.6 million to build an identical 10-metre telescope, Keck II.

Like Keck I, the new telescope will be built on the 13,600-foot Hawaiian volcano Mauna Kea by a consortium of Caltech and the University of California. To allow optical and infrared interferometry, the new telescope will be built adjacent to Keck I. Separated by a distance of 85 metres, the two telescopes will have the effective resolution of a telescope with an 85-metre mirror — an unparalleled combination of light-collecting power and spatial resolution for a ground based telescope. Keck I will be completed next year; Keck II is scheduled for completion in 1996. They will be joined on Mauna Kea later in the decade by a Japanese 8-metre infrared telescope that was approved earlier this year.

Christopher Anderson

TECHNOLOGY POLICY

Materials top critical technologies list

Washington

ANSWERING a congressional plea for the US government to take the lead in setting technology priorities, the White House last week released a list of 22 'critical' technologies.

Separated into six general categories — materials, manufacturing, information and communications, biotechnology and life sciences, aeronautics and surface transportation, and energy and environment — the list is heavy on those technologies related to economic prosperity and national security and light on basic research and exploration technologies.

Topping the list are five materials technologies, reaffirming the high priority the Administration places on that field. Earlier this year, White House Office of Science and Technology Policy (OSTP) officials previewed a large, multi-agency materials research initiative that the science office is planning for the 1993 budget (see *Nature* 350, 365; 4 April 1991).

Computers dominate the rest of the list, with seven key technologies in the category of information and communications. The White House panel's selections were: software; microelectronics and optoelectronics; high-performance computing and networking (also the subject of a new OSTP funding initiative); high-definition imaging and displays; sensors and signal processing; data storage and peripherals; and computer simulation and modelling.

Biomedical research was represented in only two technologies: applied molecular biology and medical technology.

OSTP is establishing a critical technologies institute to ensure that government research spending reflects the priorities of the list. Following a Bush Administration mandate, government support will be limited to 'generic, precompetitive' technologies, to avoid being put in the position of picking industry winners and losers.

Christopher Anderson

Correction: Gamma-Ray Observatory

LAST month's account of the launching of the Gamma-Ray Observatory (GRO) (*Nature* 350, 451; 1991) contained a concise and accurate description of a grazing incidence photon detector. Unfortunately, such devices work only for X-rays, and GRO has no such thing on board. The two large instruments, EGRET and COMPTEL, are based on scintillators. In EGRET, energetic gamma-rays create electron-positron pairs, whose energy and tracks are used to reconstruct the direction of origin of the incident gamma-ray. In COMPTEL, photons from successive Compton scatterings of a gamma-ray achieve the same purpose. □

Gulf war and sanctions

SIR — On the day Iraq invaded Kuwait, I completed my thirteenth year as an associate professor at the medical school of Kuwait University. I was one of a group of expatriates who joined the university in 1977 and with the Kuwaiti staff helped to establish the medical school.

During the past 13 years, we watched the medical school grow into a well-established scientific institution. It has always set very high standards for both students and research. This was recognized by many international institutions as well as by scientists and academic visitors.

The medical school also established a sound and well-designed postgraduate programme. This was possible because of excellent planning and the availability of up-to-date equipment and materials as well as a good infrastructure.

The medical school had held an international scientific meeting every three years and symposia and workshops were also held every year in all branches of basic and clinical sciences and the proceedings of these meetings were published. This scientific activity re-emphasizes not only the faculty's reputation for outstanding original research but also that the research is directly applicable to the needs of Kuwait and the Gulf region.

The scientific community all over the world was shocked and appalled by the looting, destruction and removal of virtually every piece of equipment and materials from scientific and educational institutions in Kuwait.

Nothing similar had ever been seen anywhere in the civilized world. It is well known that Saddam Hussein's regime is despotic and tyrannical and rules by fear. But what I cannot comprehend is the dishonesty and lack of scientific decency of the Iraqi professors who were personally involved in the systematic dismantling of the various scientific institutions around Kuwait (see *Nature* **347**, 420; 1990 & **349**, 450; 1991).

I strongly believe that the scientific community should impose sanctions on scientific contributions from Iraq for some time to come, to force Iraq to pay for the equipment that was removed from Kuwait.

ABDALLA O. ELKHAWAD

University of Southampton,
Department of Physiology &
Pharmacology,
Bassett Crescent East,
Southampton SO9 3TU, UK

SIR — While it seems unlikely, from the evidence reported, that the Iraqi 'baby milk' factory was a biological weapons assembly plant (see *Nature* **350**, 117; 1991), use of the factory for the production of the bacteriological media required for the production of such weapons is a very real possibility. Peptone powders and more complex formulations could be easily produced and shipped

to a separate facility for cultivation of the microorganisms used in the manufacture of biological weapons. The apparent lack of microorganisms or their products at the Iraqi site does not exclude the possibility that the factory was indirectly involved in the production of biological weapons.

SCOTT D. COBOURN

Center for Bioorganic Chemistry,
Research Triangle Institute,
Research Triangle Park,
North Carolina 27709, USA

Infectious paradigm

SIR — Walter Gilbert's vision (*Nature* **349**, 99; 1991) that the paradigm shift provoked by the human genome project may free small science from DNA technology is convincing, but is there not a danger that the same paradigm will be applied to small science itself? So much can be seen in the criteria used in the determination of research grant applications — the certainty of a promised result, and of an advertised timetable, for example.

What is happening is that biologists are muddling together two activities which, in the physical sciences, are separate — scientific and engineering research. Both demand ingenuity and perseverance, but they are different.

Research in physics, striving for generalizations, reflects the curiosity of individuals, cannot be totally planned and does not guarantee interesting results. Engineering research, by contrast, seeks information specific to one group of phenomena that can then be put to practical use. The work is varied — the thermal conductivity of ceramics or the execution time of algorithms — and usually involves social or even corporate interests, but the methods are known, the timetables calculable and the outcomes foreseeable.

Traditionally, that difference has been reflected in the practice of funding agencies. Engineering projects have been awarded contracts, science research projects have won grants.

Medicine, the application of biological knowledge to the preservation of human health, is biology's analogue of engineering. None of us disputes that the genome project will benefit medicine, but we should not lightly assume that all of the coming decade's addition to our understanding of biology will be a by-product of molecular medicine. And we should beware of unwittingly subjecting biologists to selection pressures that only engineering research can meet.

DOUGLAS E. BRASH

Department of Therapeutic Radiology,
Yale University,
333 Cedar Street,
New Haven, Connecticut 06510, USA

British science

SIR — Terence Kealey (*Nature* **350**, 370; 1991) raises two interesting issues about the structure and performance of British science, both of which relate to studies undertaken by the Science and Engineering Policy Studies Unit (SEPSU) in recent years.

One issue concerns the interpretation of bibliometric data, on which some of Kealey's comments are misleading. First, the Institute for Scientific Information (ISI) did indeed increase its journal coverage by nearly 40 per cent between 1975 and 1982, but one cannot infer from this that the global population of journals also increased by 40 per cent, still less that Britain "increased its numbers of papers by some 30 per cent". The 40 per cent figure is primarily a statement about ISI, not about world output. Second, we at SEPSU showed in our original study that both the 1973 fixed journal set and a larger set defined in 1981 give the same figure — 8.3 per cent for the UK share of world output in 1982. We did not 'ignore' the journals established after 1973: we demonstrated that they did not materially alter our conclusions. Third, to the extent that research is a competitive business, market share matters as well as absolute volume.

The second and more important issue is the extensive change to the structure of British science that has occurred over the past 10–15 years. Kealey very rightly draws attention to the 'privatization' of British science. One consequence of this has been a switch from long-term to project funding and, with it, the rapid growth in the number of short-term academic staff that Kealey mentions. However, it is unsafe to draw too many comforting conclusions from this development: nearly two-thirds of these short-term academic staff are below PhD status and would not normally be regarded as independent researchers, while the remainder have uncertain career prospects. Moreover, while it is certainly true that industrial and charitable funding of university scientific research has doubled in recent years, it still constitutes only 35 per cent of total external funding of university research, or 18 per cent of UFC + external funding.

Drawing in part on earlier SEPSU work, the Royal Society is now conducting a major inquiry into long-term issues in science policy, to think through the implications of the above, and other, structural changes and to identify some of the policies that will be needed to sustain the health of British science into the next century (see *Nature* **349**, 183; 1991). Anyone wishing to contribute to the inquiry is invited to write to the president, Sir Michael Atiyah (The Royal Society, 6 Carlton House Terrace, London SW1Y 5AG).

PETER COLLINS

Science and Engineering
Policy Studies Unit,
6 Carlton House Terrace,
London SW1Y 5AG, UK

Wittgenstein and reality

SIR — It is quite understandable that a gung-ho positivist, who unquestioningly accepts what he sees as reality, is annoyed by someone such as Wittgenstein and wants to rubbish him. However, the grounds selected by J. R. Smythies (*Nature* 350, 9; 1991) who says that Wittgenstein was talking 'schizophrenese' — a form of speech disorder characterized by the fact that the meaning of a schizophrenic proposition is not wholly contained within it — will not do.

Of course, it would be easy to reject anything we do not understand as a distorted proposition or a non-proposition, and this is just what Wittgenstein, in common with the logical positivists, tried to do in the *Tractatus*. However, as Wittgenstein's later work (especially the *Philosophical Investigations*) shows, no proposition or set of propositions wholly contains their meaning. An understanding of the meaning of a proposition always depends upon other, external, factors. We understand one another — for example, I understand Smythies and Smythies understands his patients — because we share a common set of assumptions with which to interpret our communications. Lacking such a mutual basis we must seek it out and fail to find it before we can dismiss propositions under such a rubric as 'schizophrenese'.

It might indeed be thought that to reject an investigation of the meaning of scientific statements (which is the business of philosophy), while confidently accepting their validity, is more truly a sign of a schizophrenic (shattered) view of reality.

R. J. BIRD

Department of Applied Social Science,
Newcastle upon Tyne Polytechnic,
Newcastle upon Tyne, NE1 8ST, UK

SIR — J. R. Smythies' explanation and dismissal of Wittgenstein's philosophy is that he was a schizophrenic writing poetry in German. That necessarily implies an admirably humble confession of a difficulty in grasping some of the illuminating aphorisms, yet it is as helpful in evaluating Wittgenstein's thought as the comment on a recent esoteric French philosopher, that he died of AIDS.

Turning the tables on Smythies, one can note that his letter reveals his quasi-religious paradigmatic belief that the neurosciences are those "from which . . . our only true understanding of the mind can come". Wittgenstein saw philosophy as therapy for such misplaced confidences. Perhaps he would have thought of Goethe, Shakespeare, Cervantes, Kafka *et al.* This is no denial of the validity and achievements of scientific enquiry. There is, however, a warning that much thought, as is most especially true for example of mathematical creations, may depend on axioms utterly obvious but only to whole generations

influenced by Zeitgeist. Extrapolating from what is successful for some issues to everything is at least questionable. Is it too germanic to note that mysteries about Being remain, or to wonder whether there is something wrong in the basic formulations we use in trying to grasp how conscious experience could evolve from matter as we have to conceive it?

Wittgenstein was undoubtedly a very disturbed and distraught person but in the light of psychiatry's own difficulties with the word schizophrenia, is much achieved by a diagnosis in this context? I must, of course, confess my own insecure and heretical position. It stresses the personal politics involved in indicating the similarities or differences between us and those called schizophrenic. That is done in part to emphasize either our ability to empathize or their strangeness. Much that many of them say is to some of us instructive as they are to a significant degree defined by their apparent emotional need to stand painfully outside complicity with the common sense of the period. Perhaps it is appropriate here to quote Karl Jaspers (1946), who went a long way to define who are schizophrenics. They are those whom among other things, he said, we cannot understand (*Verstehen*) humanly so must explain (*Erklären*) scientifically. Nevertheless, he added (much nearer to my views), "extreme psychotic states offer a parable — patients see into depths which do not belong so much to their illness as to themselves as individuals with their own historical truth . . . in psychotic reality we find an abundance of content representing fundamental problems of philosophy . . . The philosopher in us cannot but be fascinated by this extraordinary reality and feel its challenge."

F. A. JENNER

University of Sheffield,
Department of Psychiatry,
Royal Hallamshire Hospital,
Glossop Road,
Sheffield S10 2JF, UK

Jaspers, K. *Allgemeine Psychopathologie* (Springer, Heidelberg, 1946).

Cancer—EMF connection

SIR — *Nature* has recently (349, 554; 1991) discussed the cancer-electromagnetic field connection. When two phenomena correlate, but no obvious direct linkage exists, it is worthwhile asking whether an indirect linkage may exist.

Cancer is well known to correlate with action of highly reactive chemicals on tissue. The presence of certain types of electrical systems can result in the production of highly reactive chemicals. Thus, for example, high-voltage transmission lines can be expected

to produce ionization of air as a result of discharges, and production of low levels of materials such as ozone and nitrogen oxides. Similarly, if electric hair-driers have motors with brushes, the sparks at the brushes could have similar results, while the high voltages used in television sets for the cathode-ray tubes could potentially do the same.

The reactive gases would have a relatively high local concentration near the device that produces them and so potentially account for the reported correlations. Could there be a correlation between the use of heavier wiring and large motors (because of starting currents)? This could again help produce the unexpected correlation.

NEVILLE J. WOOLF

University of Arizona,
Steward Observatory,
Tucson, Arizona 85721, USA

Health risks

SIR — David Lindley (*Nature* 346, 507; 1991) makes the valid suggestion that public fears about health risks could be mitigated if scientists presented statistical arguments more precisely so that the public understood them better. In passing, however, he appears to give credence to the popular misconception that nuclear power is 'safe' compared with not wearing a motorcycle helmet or seat belt. But in terms of unit risk of death, the converse is true. Motorcycle helmets and seat belts have not been proved to reduce individual risk of death (and probably do not, see J. G. U. Adams *Ergonomics* 31, 407–428; 1988), whereas nuclear power is known to have killed.

This example illustrates the important point that statistical arguments are futile when addressing public health fears because the objective statistical information is simply not available in matters of national policy. Where, for example is the evidence that crash helmets and seat belts 'save lives'? These unproven devices are promoted as beneficial (in the absence of evidence) and thus deflect public attention from the fact that a transport policy that encourages private road transport (responsible for 5,000 accident fatalities a year in the United Kingdom) is inherently dangerous. In similar vein, energy efficiency, although cheaper and inherently safer than nuclear power, is not national policy and thus not promoted. People certainly often appear to be 'penny-wise and pound-foolish' in their personal risk perception. But underlying this actuarial schizophrenia is the uncomfortable reality that public policy often dictates the risks the public are forced to take. The public, of course, orders its fears accordingly.

SIMON P. WOLFF

Department of Clinical Pharmacology,
UCMSM,
University College London,
5 University Street,
(The Rayne Institute),
London WC1E 6JJ, UK

Funding of molecular biology

Benno Müller-Hill

The way that molecular biology is funded in Germany is damaging the universities. German policy makers should follow the advice of Wilhelm von Humboldt, and concentrate teaching and research in the universities.

MONEY for fundamental research in molecular biology is getting tighter. Only one in ten of the grant proposals put forward today by newcomers is funded by the National Institutes of Health in the United States. Although the situation is better in Europe, the strain is noticed there too. And it is felt differently in different countries and institutions. Germany is interesting in this respect because its fundamental research is divided into two parts.

On one side are several research centres such as the Max-Planck-Institutes or the German Cancer Centre. The directors of these institutions receive ample ground support and can, in addition, apply to the government for grants. But they do not have to prove their productivity repeatedly to receive this support — it will be given to them from their appointment until their retirement. A convincing argument in favour of this procedure is that it encourages truly outstanding researchers to attack fundamental problems, which often take several years to solve. Also, only research centres can afford extremely expensive equipment.

On the other side are the universities. Most German university professors get little or no ground support for their research. That which they do receive just suffices to fund practical classes and small libraries. Thus, almost all of the money for research in most universities comes from granting agencies. The largest such agency in Germany is the Deutsche Forschungsgemeinschaft (DFG). Its budget of more than DM 1,000 million is as large as the budget of the Max-Planck-Gesellschaft (MPG), the society which funds the Max-Planck-Institutes.

This has not always been so. The number of scientists has grown exponentially since 1700, doubling about every 15 years. Thus only 1.5 per cent of the present number of scientists existed in 1900. At that time, all research in biology and chemistry in Germany was done exclusively in university institutes or in industry. State-financed research centres

did not yet exist in these two fields. This was due to the effort of the founder of Berlin University, Wilhelm von Humboldt, who had fought successfully against research in academies and for the unity of teaching and research within the universities.

This changed in 1911 when German Kaiser Wilhelm, well known for his sense of humour, used the centennial of Berlin

Deutscher Forschung in 1920, the predecessor of the present DFG. Its specific goal was to help university research. It introduced peer review and two-year grants. At the end of the Weimar Republic, about 90 per cent of its money for biology went to university institutes. As industrial money vanished, the Kaiser-Wilhelm-Institutes needed and got more and more money from the state.

Mary Evans When the Nazis came to power, they removed all Jewish scientists, but they did not abandon the remaining German ones. In the absence of notable inflation they more than doubled the budget of the Kaiser-Wilhelm-Gesellschaft from 1932 to 1940/44 and they increased the DFG grant for fundamental biology fivefold from 1932 to 1940/44. They centralized the review system of the DFG and changed the distribution of the money: between 1940 and 1944, about half of the DFG grant for biology went to Kaiser-Wilhelm-Institutes. The Kaiser-Wilhelm-Institutes clearly prospered financially during the regime of the Nazis.

After the Second World War, Germany was in shambles. The Kaiser-Wilhelm-Gesellschaft was recreated as the MPG in 1948. From this time on it relied almost exclusively on state money. The former Notgemeinschaft was recreated too as the DFG, actually under the name the Nazis had given the institution but with the old rules from the Weimar Republic.

Today both organizations accept

and spend about the same amount of money received from the state, and more than half of all German research in molecular biology is done in research centres, assuming that the number of Germans in the European Molecular Biology Organisation is representative of this.

Nobody knows how long exponential growth in the number of scientists will go on. At the moment about one per cent of the population of the United States could be called scientists (that is, they have a low science degree). But there has not been a proportional growth in the number of top-

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Wilhelm von Humboldt (1767–1835)

University to announce the foundation of research centres outside the universities with the financial support of industry. These Kaiser-Wilhelm-Institutes were small to begin with. The best of the university professors were hired as directors, who rarely had more than three postdoctoral researchers and very few graduate students. This enviable situation did not last long; the First World War began and ended in disaster for Germany.

The abysmal economic situation after the First World War was the proper occasion for the foundation of the Notgemeinschaft

rate scientists. It is an illusion that all directorships of the Max-Planck-Institutes can now be filled by people who will be as productive as Butenandt, Kuhn, Meyerhof or Warburg, to mention just a few Nobel prize winners from the first half of this century. The MPG can now attract all the best and most of the better molecular biologists.

The universities can no longer compete in Germany. There have been very few cases of university professors who have turned down offers of appointments made to them by the MPG. Most university professors are more than happy when they can exchange their poor university chair and uncertain grants for a Max-Planck position, and reverse traffic of directors from the Max-Planck-Institutes to the universities is almost unheard of. It is true that universities can get good researchers from the Max-Planck-Institutes, but the research centres offer strong incentives for outstanding researchers to stay with them as permanent members or as division heads. This situation is beginning to damage the German universities. If more and more of the better research is coming from the Max-Planck-Institutes and other such centres, then their representatives will begin to argue that their funding should increase. This will inevitably lead to the collapse or reshaping of the German universities: they will become places where undergraduates are taught by professors who do not do research, and graduates will move to the centres to do the research for their doctorates.

Is this a bad prospect? History has already informed us of the answer. The United States uses a grant system based on the German *Notgemeinschaft* or DFG to fund its flourishing research, which is conducted mainly in state or private universities. The Soviet Union has opted for the Kaiser-Wilhelm system by concentrating almost all of its research in Academy Institutes. If the Academy Institutes in Moscow are compared with the universities in Boston it is beyond question who is the winner in molecular biology: the Americans. The few seminal discoveries made at Academy Institutes never came from the ground. The directors of the Academy Institutes, with between 200 and 500 graduates to oversee, were simply unable to recognize and promote outstanding individuals in time. Of course it can be argued that the one-party system with its effective secret police and its ineffective planning undermined all sincere efforts. This is certainly true, but I suspect that the main defect is elsewhere: in the near absence of functioning universities, the monopolistic research-centre system inherently performs very badly.



The Institute for Genetics, Cologne

COMPARISON OF THREE GERMAN RESEARCH INSTITUTES

	MPIB	MPIMB	IGUC
Permanent budget in 1982 (DM million)	32.0	12.5	4.9
Grants in 1982 (DM million)	2.9	1.3	4.4
Total budget in 1982 (DM million)	34.9	13.8	9.3
Laboratory space (m ²)	20,434	5,400	2,883
Doctoral dissertations	72	43	33
Articles	629.2	370.0	168.0
Cost of articles (DM thousand)	55.5	37.3	55.6
Citations	7,016.2	3,906.9	3,839.0
Citations per article	11.1	10.6	22.8
Costs of citations (DM thousand)	5.0	3.5	2.4

Indicators of the scientific productivity of the Max-Planck-Institute for Biochemistry in Munich-Martinsried (MPIB), the Max-Planck-Institute for Molecular Genetics in Berlin (MPIMG) and the Institute for Genetics of the University of Cologne (IGUC), 1980-1984. Yearly budgets are as follows. MPIB and MPIMG for 1982, data provided by the Generalverwaltung of the MPG. The sum does not include the costs of administration. Grants provided by DFG through Normalverfahren are not included either. IGUC for 1982, data provided by the administration of Cologne University. The sum includes the salaries of all employed, teaching costs, research costs, electricity, heat, maintenance, and so on. It does not include central administration or other central costs. All DFG grants and grants of the Bundesministerium für Forschung und Technologie (BMFT) are noted. Number of doctoral dissertations finished between January 1980 and December 1984 according to *Jahrbuch der MPG* and information of the Dean's office of Cologne University. Articles: the yearbooks and reprint collections of the various units of the various institutes were comprehensively consulted. All articles including short notes, abstracts, articles in books, education reviews which were cited at least once by other researchers were counted. These articles were weighted in the following manner. When two groups collaborated they were counted as 0.5, when three groups collaborated as 0.33, and so on. Citations: the *Citation Index* of 1980-1984 and 1985, 1986, 1987 and 1988 were consulted. Misspellings and umlauts were considered. Citations of papers were collected for the year of appearance and for four further years. These citations were weighted in the following way. When two groups collaborated, and the Munich, Berlin or Cologne group provided the first author, the number of citations was multiplied by 0.66, when these laboratories did not provide the first author, the citations were multiplied by 0.34. When three groups collaborated, the citations were multiplied by 0.5 and 0.25 respectively.

The failure of the Soviet research centres is striking, but are not the Max-Planck-Institutes much more productive than comparative university institutes? To answer this question, my collaborator H. Herbertz and I have compared and evaluated Max-Planck and university institutes which are active in research in molecular biology or biochemistry. The table compares two Max-Planck-Institutes and my own university institute. My university institute — it was at the time the largest and richest in Germany — did at least as well if not better than each of the two Max-Planck-Institutes in terms of the cost of articles and citations or the number of citations per article. The data presented indicate that research at a comparable university institute was effective when it had full support through the competitive grant system of the DFG.

The universities are productive in that they train the future experts for industry and the research centres. Industry and the state will suffer if further growth of research centres catalyses the destruction of the universities. To prevent this, the research centres could support outstanding university institutes by grants. German policy makers should follow the advice of Humboldt and concentrate teaching and research in the universities. Moreover, they should increase strongly and steadily funding of the universities and their granting agency, the DFG, and slow down or even freeze further growth of the research centres. □

Benno Müller-Hill is in the *Institut für Genetik der Universität zu Köln*, Weyertal 121, D-5000 Köln 41, Germany.

Another mountain from a molehill

Stealing ideas is bad for research and the reputation of science, but so are unfounded allegations that there has been a theft when there is none, now put in circulation by a materials scientist from Pennsylvania.

PLAGIARISM is the most serious of the known crimes against scholarship. Not only does it confuse the record by making it impossible to tell who exactly said what, but it amounts to the literal theft of another's words, thereby depriving the victim not merely of the credit for the content of the stolen words but of whatever thought and imagination they embody. It is a shabby business.

So what about the appropriation of others' ideas? That is also shabby, whence the convention that authors should be meticulous in their references to the antecedents of their research. By failing to refer to a significant earlier article, an author misleads his readers about the genesis of what he or she has to say and also captures an undue proportion of the credit for what is new. Oversights of this kind, deliberate or otherwise, are the most common cause of disputes about priority. People will even hotly claim precedence for ideas recognized to be hare-brained, but even that is understandable.

In short, the omission of crucial references is reprehensible, may amount to the theft of other people's ideas and, as far as possible, should be stopped. The excuse that 'there is no copyright in ideas' will not wash. But there are obvious difficulties. What about, for example, when two groups of people make more or less the same discovery at more or less the same time and submit accounts for publication separately? One may accuse the other of having picked up crucial information at a meeting, or may even suspect that referees have spilled the beans.

Obviously, some oversights can also be innocently explained; an author may not have known of earlier related work. A recent article on this page (*Nature* 349, 363; 1991) which drew attention to a recent reassessment of the importance of the expansion of the Universe and its age in the resolution of Olbers' paradox has, for example, already drawn the comment that a similar conclusion was reached in a contribution to *Nature* nearly 30 years ago.

Such incidents, while still regrettable, are likely to become more common. However the bibliographic databanks are embellished, there is a high risk that anticipatory work will be overlooked, at least in fields not constantly reworked by the textbook and monograph writers. As the record on which contemporary research is built stretches further back, such occurrences will multiply, at least until the bibliographers learn to index their material by concept as well as content.

Most people accept these difficulties — but not, it appears, Professor D. M. ('Rus-

tum') Roy of the Materials Research Laboratory at the Pennsylvania State University, who is well-known as a proselytizer of materials science. Roy is now deeply offended that *Nature* has published an article that fails to refer to two earlier publications (also in *Nature*) in 1971 and 1974 with which he had been associated. Ordinarily, the first step in such a saga is a letter of protest to the offending journal, which, fair play, Roy sent on 13 March, but intemperately. Most of what follows is concerned with what has happened since.

The article about which Roy protests is that by Patricia A. Bianconi, Jun Lin and Angela R. Strzelecki (*Nature* 315, 349; 24 January 1991). Piquantly, it may be thought, the authors are also at Pennsylvania State University, but in the chemistry department. They point out the potential importance of synthetic materials in which inorganic constituents are precipitated within an organic matrix and describe the precipitation of CdS crystals within a synthetic organic matrix consisting of polyethylene-oxide.

Nobody would mistake the article for a claim to have made a usable material; its chief interest, as explained by Stephen Mann in an accompanying article (*Nature* 349, 285; 1991) is that the authors have been able to find conditions under which apparently amorphous CdS makes a phase transition to a cubic form, suggestive of how the mineral components in materials such as bone and teeth are often orientated and otherwise disposed so as to enhance their links with the underlying matrix. Appropriately, but, as events have shown, perhaps unfortunately, the authors use the terms 'biomimetic' and 'biocomposite'.

Roy has taken up the cudgels on behalf of two earlier articles. One describes the formation of porous bone-like structures by using a resin to make a cast of the pores in natural bony material (the skeleton of the echinoderm known as the slate-pencil urchin) in which ceramic materials are then precipitated (J. N. Weber *et al.* *Nature* 233, 337; 1971). The other describes a technique for replacing the carbonate of coral by phosphate to form harder apatite (D. M. Roy & S. K. Linnehan *Nature* 247, 220; 1974). Both principles have since been used in the fabrication of synthetic composites.

Roy's cudgels have been exercised energetically. He has, for example, distributed widely an undated memorandum "in the interests of maintaining the integrity of the modern science enterprise" which says that, while this is not a case of "cheating, fraud,

malice, etc.", *Nature* and the Pennsylvania State University have acted jointly in the overselling of science, which is being advertised "like cigarettes using inaccurate misleading exaggerations". The same document complains that *Nature* made things worse by inviting Stephen Mann to comment on Bianconi *et al.*

This would ordinarily have been but a storm in a teacup had not the document fallen into the hands of a news magazine to which Roy confided that he knew from a telephone conversation that the editor of *Nature*, correctly identified by first name, was annoyed ("pissed off" was the verb he chose) with Mann for having written an egregious puff, but that Roy did not know whether the outcome would be a scandal or a "cover-up". Yet in the only conversation on this or (if memory serves) any other subject, Roy said merely "Why would I say a thing like that?" It is mystifying, to say the least.

So is the feasibility of the remedies Roy advocates. He would, for example, require all publications to include an explicit statement of how the authors had searched the literature, that journals should be compelled to publish amendments to lists of references and that grant-making agencies should appoint ombudsmen to assess the validity of claims to novelty. (One little irony is that the two earlier articles do not use the word 'biomimetic', on which a literature search would have been conducted, but which may not have been invented in the 1970s.) There was even at one stage talk of having this apparatus given teeth (in the United States) by an act of Congress.

All of us will appreciate, all too often from first-hand experience, the acuteness of the sense of neglect evoked by recognizing our own thoughts converted by others into words, and the ferocity of the emotions that can be aroused when others occupy intellectual territory we regard as ours. (This, no doubt, explains Roy's assertion that using the word 'biomaterials' in connection with the precipitation of CdS in a polymer gel is tantamount to 'poisoning the language of science'.) Yet even by those tests, Roy's reaction to Bianconi is no less exaggerated than the supposed overselling of which he complains.

Consciously appropriating the ideas of others without acknowledgement is bad for science, but so is making a needless fuss about empty issues. Indeed, gratuitous unseemliness may be just as damaging, suggesting to others that researchers are incorrigibly petty and quarrelsome. **John Maddox**

Limit to greenhouse warming?

Andrew J. Heymsfield and Larry M. Miloshevich

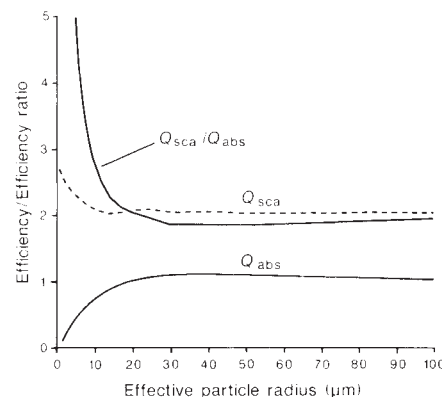
THE greenhouse effect, a net atmospheric warming attributable primarily to the absorption of infrared (longwave) radiation by water vapour and clouds, may be limited in magnitude in the tropics by a feedback mechanism involving sea surface temperature (SST) and cirrus ice clouds, as proposed by V. Ramanathan and W. Collins elsewhere in this issue (*Nature* **351**, 27–32; 1991). They find that when the SST exceeds about 300 K, vigorous, deep convective clouds rise to high altitudes and produce extensive cirrus ice clouds, reducing solar heating and limiting the SST to, at most, 305 K. In the absence of some such means to limit solar surface heating, the rising SST and consequent build up of water vapour in the atmosphere would result in a runaway greenhouse effect. This mechanism may have important implications for global climate change.

In the absence of an atmosphere, the fundamental factors controlling the planetary heat balance are the solar (shortwave) energy absorbed at the surface (the incident energy minus the energy reflected, or surface albedo) and the longwave energy emitted by the surface back to space. The atmosphere contributes to the heat balance through the absorption of shortwave radiation by atmospheric gases, addition of a cloud albedo to the surface albedo, and net longwave absorption by cloud drops and ice particles and by greenhouse gases (particularly water vapour, carbon dioxide and methane). The greenhouse effect is caused by the longwave absorption by atmospheric water vapour (and other greenhouse gases), and the net longwave absorption in clouds.

Ramanathan and Collins determine the upper limit of SST by correlating satellite radiation data with SST measurements over a period which includes the 1987 El Niño event, an occasional climate fluctuation affecting the southern Pacific area. They then establish values for the various radiative contributions to the Earth's heat budget.

The satellite data show a pronounced increase in the greenhouse effect as SST increases above 300 K. The authors find that the longwave absorption by clouds is the dominant contribution to the greenhouse effect, and interpret this observation as indicating an increase in cloud extent and optical thickness as SST increases. The data also show that cloud albedo increases as SST increases, thus the clouds that trap more longwave radiation are also more reflective to shortwave radiation. Furthermore, the data show that shortwave reflection increases with increasing SST at a faster rate than the corresponding increase in greenhouse warming. Eventually the increased cloud production provides sufficient shortwave blocking to inhibit further increases in SST, and prevents a runaway greenhouse effect.

Using the observed values for the rates of change of atmospheric and cloud contributions to the radiative heat budget, Ramanathan and Collins determine that the maximum possible tropical SST is between 303 and 305 K, in agreement with the maximum value of 304 K measured during the 1987 El Niño. The range of values given for the maximum SST is largely a result of the bounding values, taken to be 0.8 and 1.0,



The radiative effect of changing ice particle size: the 'Mie' scattering efficiency, calculated for spherical particles, is shown for shortwave reflections (at a wavelength of 0.65 μm; Q_{sca}) and for longwave absorption (at 10 μm wavelength; Q_{abs}). The ratio of these (Q_{sca}/Q_{abs}) is also shown (solid line). (Calculations by Patrick Minnis, Langley Research Center.) The effect of size on the radiative properties of ice particles could contribute to the feedback observed by Ramanathan and Collins, but *in situ* measurements of real nonspherical ice particles in clouds will be needed to clarify the details. This example, using one particular set of wavelengths, can only be taken as an indication of possible particle size effects on cloud radiative properties, as absorption efficiencies depend strongly on wavelength and multiple-scattering of shortwave radiation may also be important.

used to describe the fraction of greenhouse warming that is advected out of the tropics and is thus unavailable to affect the SST directly. These high values for the fraction of advected heat suggest that the feedback mechanism, which occurs in the warmest regions of the planet, may have consequences for the planet as a whole.

The mechanism that the authors propose to explain the data involves the effects of solar heating of the surface and resulting convection on the balance between greenhouse warming and cloud reflectivity. When the SST is below 300 K, the latent energy in parcels of moist air near the sea surface is insufficient to produce deep convection, but, when it exceeds 300 K, parcels have sufficient latent energy to rise to the upper troposphere

as cumulonimbus clouds. At the cumulonimbus cloudtops, ice crystals form cirrus clouds spreading thousands of kilometres downwind, dominating radiative processes in the tropics when the feedback mechanism is operating. Because the optical depth of the cirrus clouds and their areal extent increase in response to increasing SST, the cloud albedo also increases, thus slowing and eventually preventing further increases in both SST and the associated greenhouse warming.

The authors suggest that increases in cirrus extent and optical thickness are responsible for the enhanced cloud albedo. It seems likely that changes in the distribution of ice particle sizes contribute to enhanced cloud reflectivity. The albedo of a cloud depends ultimately on the scattering behaviour of individual ice particles (the phase function, dependent on particle size and shape), integrated over both the particle size distribution and the path length through the cloud. Studies of stratiform clouds (A. J. H. & C.M.R. Platt *J. atmos. sci.* **41**, 846–855; 1984; see also the special 'FIRE' issue *Mon. Weather Rev.* **118**, 2259–2446; 1990) suggest that ice particles tend to be smaller at lower temperatures, with diameters of a millimetre or more near 0 °C, but only tens of microns at the temperatures characteristic of the tropical tropopause. This is a result of changes in diverse factors such as cloud water vapour content, cloud thickness, particle terminal velocity and vertical air velocity. These observations in stratiform clouds may not apply to the convective regions of the cumulonimbus; but, fallout of the larger particles with increasing distance downwind of the convective towers will leave cirrus clouds containing smaller ice crystals.

The effect of the size of ice particles on their radiative properties depends on their shortwave scattering efficiency (Q_{sca}) and their longwave absorption efficiency (Q_{abs}) for a given cloud volume and particle size distribution. The figure shows that, at least for spherical ice particles, the ratio $Q_{sca}/Q_{abs} \approx 2$ for particles larger than 25 μm, but the ratio increases significantly as their size decreases below 25 μm. As the SST increases, production of colder cirrus clouds may contain smaller ice particles which may lead to enhanced shortwave reflection relative to longwave absorption, in addition to the enhanced shortwave reflection caused by increases in cirrus ice mass alone.

It is unclear what relevance the feedback mechanisms proposed by Ramanathan and Collins have to the issue of CO₂ greenhouse warming: the climate's response to increases in atmospheric CO₂ could be very different from the negative-feedback response to solar heating that the authors describe. The source of the warming is the obvious difference. Whereas solar radiation directly heats the ocean surface, giving rise to the SST-limiting feedback, CO₂ produces warming in the overlying atmosphere through trapping of surface-emitted longwave radiation. As

most upper-tropospheric heating in the tropics is advected out of the region, CO₂ warming in the atmosphere should have much less of an effect on limiting the SST than does direct solar heating, although the relationship between atmospheric warming and SST is a subject of controversy. In addition, atmos-

LIQUID CRYSTALS

Shedding light on alignment

Cliff Jones and Sally Day

ORGANIC materials have enormous potential for use in electro-optic and opto-optic devices. Nowhere is this more evident than with liquid crystals, which are already familiar in the displays of digital watches and laptop computers. These devices are electro-optic: applying an electric field reorients the liquid crystal molecules, thereby altering the device's optical properties. New effects and applications are being sought, and one particularly exciting possibility is that of controlling the molecular orientation using incident light. Such an effect is described by Gibbons *et al.* on page 49 of this issue¹, who show how to manipulate the orientation of a nematic liquid crystal by irradiating the surface alignment layer with polarized light.

Vital to most liquid crystal devices are surface alignment agents which define the molecular orientation of the device in the unswitched state. Numerous surfactants are available, but the usual commercial technique is to coat the inner surfaces of the glass cell containing the liquid crystals with a polymer such as polyimide. This causes the rod-like liquid crystal molecules to lie parallel to the surface. Rubbing the surface with a nylon cloth then imparts a preferred alignment for the liquid crystal molecules once the cell is filled. For example, in the common twisted-nematic device, the surfaces are rubbed at 90° to each other, so that the liquid crystal molecules twist through this angle from one surface to the other. Light is transmitted through the liquid crystals when viewed between crossed polarizers, and the cell appears light. Applying a small voltage (typically about a volt) reorients the molecules, removing the twisted structure so that the device is dark. In this manner the twisted-nematic cell acts as a simple electro-optic shutter.

Gibbons *et al.* use a novel alignment layer that is a mixture of polyimide and an azo-dye. Anisotropic dyes of this type are often used in liquid crystal applications because they align parallel to the liquid crystal molecules, providing anisotropic absorption of incident light and allowing displays incorporating a single polarizer. Gibbons *et al.* constructed a cell in which one surface is coated with polyimide and the other with the dye-doped polyimide, and the plates are aligned with rubbing directions parallel. The liquid crystal is then uniform and behaves optically as a birefringent slab (so that, for

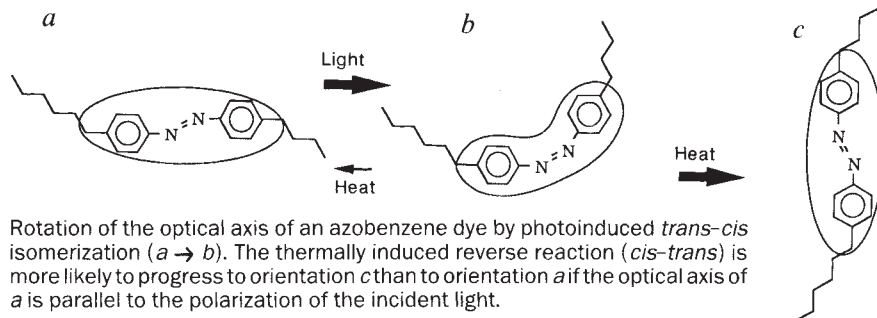
example, it would appear dark between crossed polarizers). Incident light from an argon-ion laser, polarized parallel to the initial direction of alignment, causes the liquid crystal molecules at the dye-doped polyimide surface to reorientate through 90°, forming a twisted-nematic structure which transmits light between crossed polarizers.

Clearly, this effect is due to the presence of the dye in the alignment layer. Similar effects have been observed in the bulk of azo-dye-doped monomeric² and side-chain polymeric³ liquid crystals. Anderle *et al.*⁴ suggest a mechanism for the photoinduced reorientation of the liquid crystal based on *trans-cis* isomerization of the azo-dye, changing the molecules' shape (see figure). Illuminating the dye with light of wavelength close to its absorption maximum induces the *trans-cis* reaction, with highest reaction probability for molecules parallel to the plane of polarization of the incident light because of the anisotropy of the dye's absorp-

sponse of the reorientation process.

Photoinduced isomerism may also lead to the surface alignment effects described by Gibbons *et al.* Two mechanisms have been proposed for the alignment of liquid crystals on rubbed polymers. One involves a long-range interaction due to surface micro-grooves caused by the rubbing. The elastic energy of the liquid crystal is then minimized if the molecules lie along the grooves. Alternatively, the alignment mechanism may be a short range interaction between the liquid crystal molecules and the polymer surfactant. It is possible to distinguish between these mechanisms through measurement of the optical second-harmonic generation (SHG, a frequency-doubling effect) of the noncentrosymmetric monolayer at a surface⁵. The liquid-crystal/rubbed-polymer interface has been studied using this powerful technique by Chen *et al.*⁶, who show, from the symmetry and polarization dependence of the SHG signal, a preferential, polar orientation of the surface layer of crystal along the rubbing direction of polyimide. Other alignment layers do not show such short-range, polar interaction and it is presumed that the alignment mechanism is due solely to the longer range elastic forces in the liquid crystal bulk. Although a dyed polyimide surfactant has not been studied by this method, the results of Gibbons *et al.* indicate that the liquid crystal molecules interact even more strongly with the dye than the polyimide.

Interesting new uses of liquid crystals may be emerging that take advantage of the fine,



Rotation of the optical axis of an azobenzene dye by photoinduced *trans-cis* isomerization (*a* → *b*). The thermally induced reverse reaction (*cis-trans*) is more likely to progress to orientation *c* than to orientation *a* if the optical axis of *a* is parallel to the polarization of the incident light.

ance. The thermally excited reverse reaction, *cis-trans*, may occur by either of two routes but the final state is most likely to be perpendicular to the polarization of the incident light because of the continued excitation of the molecules in that direction.

A liquid crystal mixed with the dye will also become reoriented under these circumstances, owing to the molecular interaction between the dye and the liquid crystal, and therefore the optical axis of the sample will rotate to be perpendicular to the incident polarized light. Eich and Wendorff³ have begun to demonstrate the considerable potential for these effects in liquid crystal polymers as a medium of erasable optical storage, producing diffraction gratings and holograms. The possibility of real-time holography for display purposes in monomeric liquid crystals² is however limited by the slow re-

adaptive control of the orientation of liquid crystals using light instead of the conventional fixed electrode structures. But the mechanisms involved must be understood more fully in order to optimize the efficiency, response times and optical characteristics needed for practical devices. □

Cliff Jones and Sally Day are at the Electronics Division of the Defence Research Agency, RSRE, St Andrews Road, Malvern, Worcestershire WR14 3PS, UK.

- Gibbons, W. M., Shannon, P. J., Sun, S. & Swetlin, B. J. *Nature* **351**, 49-50 (1991).
- Anderle, K. *et al.* Proc. 20th Freiburger Arbeitstagung Flüssigkristalle, P22 (Fraunhofer Inst., Freiburg, 1991).
- Eich, M. & Wendorff, J. H. *J. opt. Soc. Am. B*, **7**, 1428-1436 (1990).
- Anderle, K. *et al.* *Liquid Crystals*, **9**, 691-699 (1991).
- Shen, Y. R. *Nature* **337**, 519-525 (1989).
- Chen, W., Feller, M. B. & Shen, Y. R. *Phys. Rev. Lett* **63**, 2665-2668 (1989).

Twinkle, twinkle

GRAVITATIONAL lensing may assist those searching for brown dwarfs, substellar objects that could account for much of the mass of our Galaxy. Because they are too small to 'burn' by nuclear fusion, brown dwarfs cannot be seen by their starlight, which is why they remain merely hypothetical. B. Paczyński reminds readers of the *Astrophysical Journal* (371, L63-L68; 1991) that the central bulge of our Galaxy is dense with slowly moving stars, and that as they pass behind the nearer stars in the galactic disk, gravity will focus their light to give us a brightened image. He points out that brown dwarfs could also focus background stars, though the brightening would not last so long (3-20 days instead of 1-4 weeks). In work in the press with *Astrophysical Journal*, Paczyński and S. Mao show that the same principle could be used to discover binary stars and remote planetary systems.

Bone hard

BONES are found far more often than soft tissue at archaeological sites, and offer a potentially huge harvest of DNA for the polymerase chain reaction — but only if the spectre of contamination can be exorcised. Hence the apparently inexhaustible capacity for taking pains shown by E. Hagelberg and J. B. Clegg in the isolation of mitochondrial DNA from a variety of bone samples (*Proc. R. Soc. B* 244, 45-50; 1991). A pig bone from the larder of the *Mary Rose*, the ill-fated, sixteenth-century flagship of the English navy, for example, yielded DNA that is unequivocally porcine. Extracting DNA from bone, then, is not a problem, and the researchers recommend a stringent protocol to minimize spurious results from the exogenous DNA invariably associated with an archaeological sample.

Change for the worse

STUDIES reported by K. Hsiao *et al.* strengthen the case that damaged prion protein is at the root of several neurological diseases (*New Engl. J. Med.* 324, 1091-1097; 1991). Some communities have an unusually high incidence of Creutzfeldt-Jakob disease, one such disorder, which has prompted several explanations of its cause, including the suggestion that consumption of the Mediterranean delicacy of sheep's eyes and brains is responsible. Hsiao and colleagues now confirm that an excess of the disease among Libyan Jews is more a result of genetics than of culinary taste: the disease occurs consistently in people in which a single amino-acid change has occurred within the prion protein, and the inherited genetic mutation seems to show complete phenotypic dominance in this population. The protein's normal function remains unknown, however.

TFIIB or not TFIIB?

Phillip A. Sharp

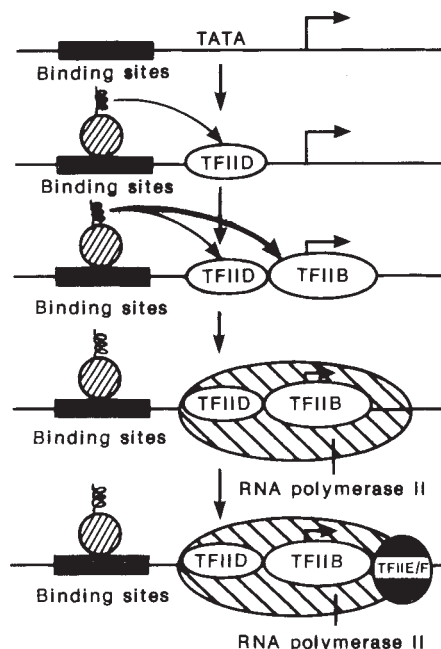
LITTLE has been established about the mechanism by which sequence-specific DNA transcription factors stimulate initiation of transcription by RNA polymerase II (B). Two series of experiments now offer hope that this record will change. First, it is likely that transcriptional activation *in vitro* may be detectable only if the basal transcription reaction is repressed by chromatin, or more specifically by components of chromatin such as histone H1 (refs 1 and 2). Second, the major activation step in initiating transcription *in vitro* is probably not, as previously proposed, the template binding of the TFIID (TATA binding) factor, but the subsequent binding of the TFIIB factor to the TFIID-template complex³. This step forms the platform for the binding of RNA polymerase II and the TFIIE/F factor.

General transcription factors activate a basal reaction at the promoter site resulting in the positioning of the polymerase, and initiation. A distressing number of factors have been described as being important in this process. The simplest scheme is based on the assembly of template complexes which can be resolved by electrophoresis in native gels (see figure)⁴. The hallmark of this scheme is the ordered addition of factors where TFIID binding permits the association of TFIIA and TFIIB. Only after binding of TFIIB does the polymerase associate and the binding of polymerase is necessary for the association of TFIIE/F. This fraction is probably a mixture of two activities where the TFIIF component remains associated with the polymerase complex. Ideally, the rate of initiation of transcription reflects the rate-limited step in the assembly of this basal reaction.

Several sequence-specific transcription factors can bind to sites both proximal and distal to promoters and stimulate the rate of initiation of transcription. Because initiation requires the interaction of a number of basal factors with the TATA sequence and the polymerase, transcription factors are thought to increase the rate of the basal reaction. These factors possess one or more activation signals which, in theory, should promote the efficiency of binding or function of one or more of the basal factors that are rate-limiting for initiation by polymerase. The activation signals of transcription factors have been imprecisely defined as acidic, or proline- or glutamine-rich. Acidic protein domains that activate transcription possess a secondary structure, perhaps α -helical, and are apparently universally active in all organisms^{5,6}. The acid-rich signal of the VP-16 protein of herpes simplex virus (HSV) has become the gold standard and has been analysed by mutations that change single amino acids. Both the acidic character and the conformation of this domain are important. Illustrative of the importance of con-

formation is that a single phenylalanine-to-proline alteration in the 78-amino-acid domain inactivates the signal⁷. Given the apparent simplicity and universality of the acidic activation signal, it was possible that a single universal step was accelerated by its activity.

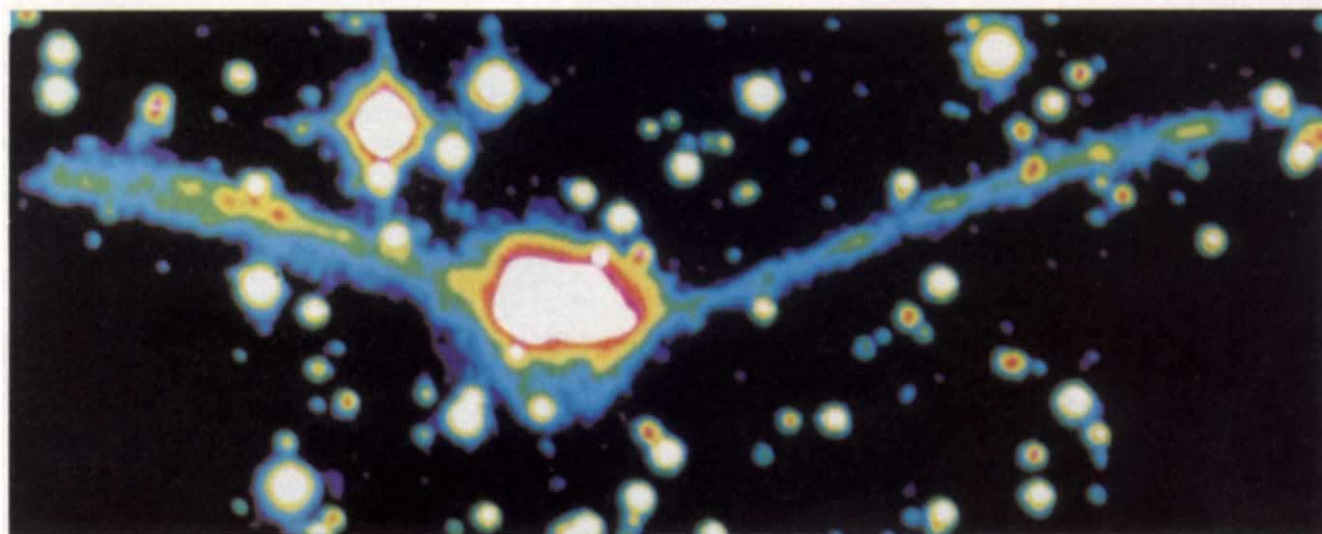
Disappointingly, addition of transcription factors with activation domains to reactions containing highly purified basal transcription factors results in only a modest 2-5-fold



A model for how an acidic activator stimulates transcription. (Adapted from refs 3 and 4.)

stimulation of transcription — not the 100-fold stimulation observed *in vivo*. Interestingly, addition of the same transcription factors to a reaction containing a total nuclear extract could stimulate transcription 100-fold. The difference between these two results has been interpreted to mean that there is a coactivator that mediates stimulation of the basal reaction by the transcription factor (for review see ref. 8), the hypothetical coactivator being missing from the reconstituted reaction.

Alternatively, a change in the nature of the rate-limited step in initiation of transcription could explain the difference. For example, the association of a basal factor might be efficient, and thus not rate limiting, in a reaction reconstituted by purified basal factors, but inefficient, and thus rate limiting, in a reaction formed with a total nuclear extract. The stimulation by a transcription factor which promotes the binding of this basal factor will be much greater in the reaction containing the total extract. Comparison of the efficiency of transcriptional activation *in vitro* in the presence or absence of chromatin



THE SUPERANTENNAE is the apt name given to this extraordinary object 250 megaparsecs (800 million light years) from our Galaxy. The object is the product of the collision, a billion (10^9) years ago, between two giant galaxies. Identified last year by J. Melnick and I. F. Mirabel (*Astr. Astrophys.* **231**, L19–L22; 1990), it is now shown by Mirabel and colleagues (*Astr. Astrophys.* **243**, 367–372; 1991) to be a massively scaled-up version of the more familiar Antennae,

after which it is named. From tip to tip, the tails extend across 350 kiloparsecs (our Galaxy has a diameter of 30 kpc). The tails are the result of the tidal interactions between the two rotating galaxies, and are ribbon-like strips of gas pulled from the progenitors' original peripheral disks. The bright knots along them are regions where hydrogen gas has condensed to trigger new outbursts of star formation. The apparently equal spacing between the knots is a puzzle that may

be solved by further observations using the European Southern Observatory's 3.6-metre telescope later this month. Beyond the tip of the southern tail (right-hand end) seems to be a knot as large as a dwarf galaxy. Mirabel comments that our Galaxy and our nearest neighbour, Andromeda, (comparable to the merging galaxies in the Superantennae) are approaching one another rapidly; but it will be several billion years before we know if we are to suffer a similar fate.

proteins suggests that the presence of components of chromatin may change the rate-limited step in initiation, and thus make the reaction more responsive to activator signals.

Workman and colleagues¹ have shown that reconstitution of chromatin on a template *in vitro* strongly suppresses the subsequent activity of the template for transcription with basal factors. However, if the reconstitution of chromatin is carried out in the presence of a transcription factor possessing an acidic activation signal, the basal reaction is efficient following addition of factors TFIID, TFIIB, polymerase II and TFIIE/F. In the last protocol, there was a 50-fold increase in transcription over the suppressed basal levels, dependent upon the presence of the transcription factor. The acidic domain of the transcription factor is essential for this activation of transcription following reconstitution of chromatin: if template is reconstituted in the presence of a mutant factor consisting of a sequence-specific DNA-binding domain, lacking its acidic activation domain, the template is poorly used by polymerase. Interestingly, the activation domain is not required for the binding of TFIID in the presence of the chromatin.

The histone protein H1 which binds to the linker DNA between nucleosomes is more prevalent on chromatin of transcriptionally inactive genes than on transcriptionally active genes; in fact, addition of H1 to chromatin containing 5S RNA genes inhibits transcription of these genes by added RNA polymerase III and factors⁹. The work de-

scribed by Kadonago and colleagues² suggests that inhibition of transcription by histone H1 may be the specific aspect of chromatin responsible for suppression of transcription. Again, transcription of a template in a nuclear extract depleted of histone H1 was repressed by addition of purified H1. If the template contained binding sites for the factor, addition of a transcription factor possessing an acidic activation signal prevented repression. Thus in the presence of H1 the factors activated transcription more than 20-fold, a similar stimulation to that observed *in vivo*.

A most startling result from Lin and Green³ suggests that the acidic activator functions by facilitating the association of the TFIIB factor with the TFIID-template complex. This insight arose from a novel experimental protocol, in which the template DNA was bound, by biotin modification, to a matrix of streptavidin. This permitted the rapid isolation of template-associated complexes and determination of the particular step in the figure that was accelerated by the presence of a transcription factor. Binding of the TFIID factor to the template was not increased by the acidic activator of the transcription factor. Surprisingly, association of TFIIB with the template-TFIID complex was highly dependent upon the presence of the transcription factor. The polymerase II and TFIIE/F components associated readily following the binding of TFIIB.

How does an acidic domain potentiate the binding of TFIIB with the template? The answer is probably by direct association of

the acidic domain with TFIIB. When nuclear extracts are chromatographed over a column matrix containing an immobilized form of a protein related to VP-16, TFIIB activity is specifically retained. This association is dependent upon the integrity of the acidic domain in that the change of a phenylalanine to proline inactivates the binding of TFIIB. The finding of an interaction of TFIIB with the acidic activator domain and not TFIID appears to conflict with the earlier results of Stringer and coworkers¹⁰, who reported the binding of TFIID and VP-16 using a similar assay. But Lin and Green² found that the TFIID/VP-16 association is much more sensitive to the concentration of counter ion than the TFIIB/VP-16 association. Thus, the picture emerges of a transcription factor located in a promoter or enhancer stabilizing the association of TFIIB, and perhaps more weakly TFIID, in the basal initiation complex.

From all these results one can predict that the presence of chromatin or histone H1 makes the initiation process highly dependent upon the step promoted by the acidic signal — that is, the binding of TFIIB. This follows from the facts that (1) the presence of chromatin components suppresses the basal reaction, thereby increasing the enhancement produced by activators; and (2) the acidic activator domain of these activators recruits TFIIB to the initiation complex. The simplest explanation for this is to suppose that the chromatin components obstruct the binding of TFIIB, making it the rate-limiting step in the process and thus increasing the dependence of the initiation reaction on the

presence of an activator domain which promotes the step. This suggests that the addition of a component that binds to promoter DNA in a nonspecific fashion (that is, chromatin proteins) could under some conditions satisfy some of the assay properties suggested previously for coactivators. In a mechanistic sense, little is known about the kinetics of TFIIB binding to the TFIID-template complex. The specific footprint generated by the binding of TFIIB over the initiation site is not pronounced and is confined to one strand of the template⁴, so TFIIB may not be strongly bound to the template in the absence of an activator domain.

Final answers to most of the questions addressed by these experiments await the purification and analysis of the specific proteins that constitute TFIIB and the other basal components. The most purified preparations of TFIIB contain a protein of relative molecular mass 32,000, and specific antiserum has been prepared¹¹ which reacts with pro-

tein in the template TFIID-TFIIB complex. It is highly likely that a complementary DNA encoding this factor will be available shortly, and more refined studies of the interaction of the protein with the acidic activator and TFIID will then be possible. □

Phillip A. Sharp is in the Center for Cancer Research and the Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

1. Workman, J. L., Taylor, I. C. A. & Kingston, R. E. *Cell* **64**, 533-544 (1991).
2. Croston, G. E., Kerrigan, L. A., Lira, L. M., Marshak, D. R. & Kadonaga, J. T. *Science* **251**, 643-649 (1991).
3. Lin, Y.-S. & Green, M. R. *Cell* **64**, 971-981 (1991).
4. Buratowski, S., Hahn, S., Guarente, L. & Sharp, P. A. *Cell* **56**, 549-561 (1989).
5. Hope, I. A. & Struhl, K. *Cell* **46**, 885-894 (1986).
6. Ptashne, M. *Nature* **335**, 683-689 (1988).
7. Cress, W. D. & Triezenberg, S. J. *Science* **251**, 87-90 (1991).
8. Lewis, B. *Cell* **61**, 1161-1164 (1990).
9. Schlisel, M. S. & Brown, D. D. *Cell* **37**, 903-912 (1984).
10. Stringer, K. F., Ingles, C. J. & Greenblatt, J. *Nature* **345**, 783-786 (1990).
11. Maldonado, E., Ha, I., Corest, P., Weis, L. & Reinberg, D. *Molec. cell. Biol.* **10**, 6335-6347 (1990).

A new act to swallow

Eve Marder

On page 60 of this issue¹, Meyrand and colleagues demonstrate that the same neurons can be part of several different functional networks, some of which may only exist operationally for a short time if the behaviour they control is not continuous. This contrasts with the view that each neuron has a specific and unique role in sensory processing or motor function; for example, one might imagine that the networks that generate walking are activated during walking, but when the animal is not walking these neurons are not used. The view of the organization of the nervous system that Meyrand *et al.* now provide shows surprising similarity to one seen in neural network theory, in which individual elements can be used to construct many different functional networks.

Meyrand *et al.* studied the neural networks of the crustacean stomatogastric nervous system², which consists of four ganglia: the stomatogastric ganglion with 30 neurons, the oesophageal ganglion of 18 neurons, and the two commissural ganglia, each with several hundred neurons. When removed from the animal, the stomatogastric system retains the ability to generate several different rhythmic motor patterns. These include the pyloric rhythm (0.5-2-second period), the gastric

and oesophageal rhythms (each of 5-10-second period), the cardiac-sac rhythm (30-120-second period), and a swallowing pattern described for the first time in the new work¹. The cellular properties and circuit interactions responsible for generating the pyloric and gastric rhythms are relatively well understood because the neurons involved are relatively few in number, and are large and uniquely identifiable regardless of their firing patterns, and because the motor neurons themselves are important components of the pattern-generating networks.

For many years biologists have treated the stomatogastric nervous system as if it were divided into neat, almost independent, subsystems, each responsible for the generation of a different motor pattern. In this view, neurons participate in only one motor pattern, and are classified as 'pyloric' or 'gastric' on the basis of their anatomical projections and firing patterns. But then Hooper and Moulins³ showed that the ventricular dilator neuron could switch its firing pattern between the pyloric and the cardiac-sac rhythm. And Weimann *et al.*⁴ demonstrated that many neurons previously thought to participate exclusively in the generation of

the pyloric or gastric rhythms could fire in time with either rhythm. Furthermore, Dickinson *et al.*⁵ showed that a peptide (red pigment concentrating hormone) could elicit a novel rhythm in which gastric and cardiac-sac neurons were coordinately active. Meyrand *et al.* now show that, when the animal swallows, a pair of py-

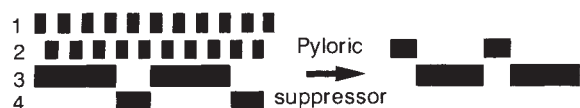
loric suppressor neurons are activated which override the synaptic and cellular events responsible for the generation of the pyloric, gastric and oesophageal rhythms, and which recruit some of the elements of these networks to make a new network to generate the swallowing motor patterns.

The pyloric suppressor neurons can be activated by stimulating sensory pathways. The neurons then act at many sites in the stomatogastric nervous system: they drive motor neurons that control the opening of the oesophageal valve for swallowing; strongly inhibit many of the neurons of the pyloric and gastric networks; and strongly excite other neurons of the pyloric, gastric and oesophageal networks. The net result is that bursts of action potentials in the pyloric suppressor neurons evoke a new motor pattern in which neurons of the pyloric, gastric and oesophageal networks fire coordinately with each other and with the neurons that operate the oesophageal valve (see figure).

It is important to remember that the stomatogastric nervous system is different from many in that the motor neurons themselves are intrinsic components of the central pattern-generating circuits. Therefore, when an identified stomatogastric ganglion motor neuron alters its activity pattern, the behaviourally relevant output and the effect of neurons that shape this output are seen simultaneously. So the new studies show that the same neuron can participate in several distinct neural networks. They also demonstrate that operational networks can be formed by synaptic and modulatory inputs when needed, but that their elements can be freed for other tasks at other times.

Some may be tempted to argue that the kinds of circuit reconfigurations that can be seen in invertebrates are not necessary in the vertebrate nervous system, with its vastly larger number of neurons. I remember when a similar argument was made about the ability of invertebrate neurons to generate bursting pacemaker and plateau potentials. Now we know that vertebrate neurons can display as rich a range of membrane properties as has been long known for invertebrate neurons. I await the time when workers on vertebrate preparations can reliably identify neurons independently of their firing patterns. We will then see the extent to which neurons in the vertebrate nervous system are used as elements in many different functional circuits. □

Eve Marder is in the Department of Biology and Center for Complex Systems, Brandeis University, Waltham, Massachusetts 02254, USA.



Block diagrams showing the activity of different neural elements. Traces 1 and 2, pyloric elements; traces 3 and 4, gastric elements. Under control conditions (left) the faster pyloric and slower gastric rhythms occur simultaneously. Pyloric suppressor activity evokes a novel pattern using elements of both the fast and slow rhythms. Other elements become inactive (traces 1 and 4).

1. Meyrand, P., Simmers, J. & Moulins, M. *Nature* **351**, 60-63 (1991).
2. Selverston, A. I. & Moulins, M. (eds) *The Crustacean Stomatogastric Nervous System* (Springer, Berlin, 1987).
3. Hooper, S. L. & Moulins, M. *Science* **244**, 1587-1589 (1989).
4. Weimann, J. M., Meyrand, P. & Marder, E. *J. Neurophysiol.* **65**, 111-122 (1991).
5. Dickinson, P. S., Mecsas, C. & Marder, E. *Nature* **344**, 155-158 (1990).

Another (and armless) army

Paul G. Bahn

ANOTHER terracotta army has been discovered in Shaanxi Province, China, in the same region as the famous one found in 1974. The new army, now described by W. Zhaojin and M. Hong (*Archéologia* 265, 4–5; 1991), is much smaller in stature, but looks to be far larger in numbers.

The discovery 17 years ago, by a farmer digging a well, of the life-size terracotta army of China's first emperor Qin Shihuangdi led eventually to this "eighth wonder of the world" becoming an attraction for tourists, who come in droves to see the 7,000 clay warriors still standing in their vault, now protected beneath an enormous hangar. Such has been the popularity of this archaeological marvel that plans were made to build a new airport, which will require a new 19-km motorway to link it with the city of Xian.

It was in the course of constructing the motorway in March last year that the new find was made. Four necropolises of the Han (206 BC–AD 220) and Tang (AD 618–907) dynasties had already been encountered in the vicinity. Tests were made in a wheatfield located 300 m south of the tomb of the wife of the emperor Liu Qi, who was known during his reign as Jing and who lived from 188 to 144 BC; he and his wife were buried at Yangling, 22 km northwest of Xian.

At a depth of 6.5 m, a small orange potsherd was found, which was soon identified as the shoulder of a statue. The whole field was then prospected, and the Archaeological Institute of Shaanxi Province sent in a team of specialists in the study of Han tombs, led by Wang Xueli, chief excavator of the famous terracotta army.

The new site proved to cover 96,000 m², the size of twelve soccer pitches. It contains an army of terracotta figures, associated with the two tombs, and housed in 24 vaults about 20 m apart and in 14 rows aligned north–south. The vaults are 4–10 m wide, and 25–291 m in length. They are therefore smaller than the Qin dynasty pits, but they cover an area about five times as large. Their different sizes may correspond to those of the army units they contain (one contains mostly cavalry, another may be the headquarters).

All the human statues are of naked men, with no arms; they are about 50 cm high. The whole body is painted an orange–red colour, with hair, eyebrows, beard and eyes coloured black. Their clothes, assumed to have been of linen or silk, have disintegrated. There are different theories about the arms. Some researchers believe they may have been held on with rods of precious metal, which were subsequently stolen. Others think it likely that they had movable wooden arms, which have disintegrated like the clothing.

The figures have graceful forms and delicately sculptured faces. Each one is different from his fellows in age and in facial expression,

which ranges from severe to smiling. In their remarkable realism they resemble the life-size figures which are about a century

swords, chisels and agricultural tools, carts, jewellery and coins — all of them, like the men and horses, one third life-size.

More than 300 statues have been uncovered so far, but it is certain that thousands remain buried; estimates of the total vary from 10,000 to a million, and excavation will continue for two to three years. Meanwhile,

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Minute men of China's new model army (picture courtesy *Archéologia*).

older. But it is the differences that are the most striking: not only are the Qin figures life-size, but their arms and their uniforms form an integral part of the statues. Also, the Qin figures are hollow, whereas the new discoveries are solid.

The new army is accompanied by weapons of copper and iron, arrowheads, spears,

the motorway project goes ahead, though now a bridge, 320 m long, will be constructed over the site, and an underground museum is being planned to enable the inevitable tourists to see the new army. □

Paul G. Bahn, 428 Anlaby Road, Hull HU3 6QP, UK, is a freelance writer on archaeology.

REPRODUCTIVE BIOLOGY

Do sperm find eggs attractive?

John Aitken

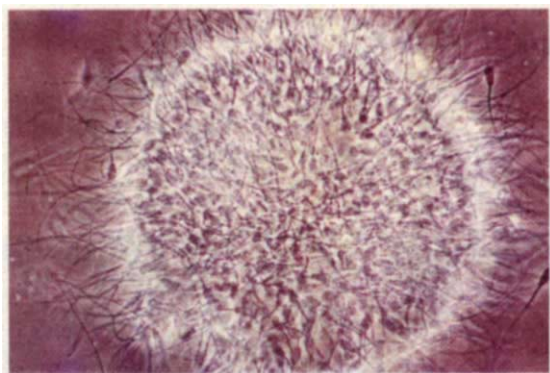
STUDIES described by D. Ralt and colleagues in *Proceedings of the National Academy of Sciences*¹ suggest that the conventional wisdom about how male and female gametes make contact in the reproductive system of the human female is wrong. The authors provide evidence for the existence of a chemotactic factor that attracts spermatozoa to the egg, thereby opening up an entirely new perspective in reproductive biology.

In the past, the best evidence for sperm chemotaxis has come from invertebrate species, such as sea urchins, that engage in external fertilization. These species shed their gametes into the ocean, necessitating the evolution of intricate species-specific mechanisms to ensure that the spermatozoa not only find eggs, but eggs from the right species. By contrast, spermatozoa in species exhibiting internal fertilization, in which the sperm are released into the confines of the female reproductive tract, were thought to

find their quarry by behaving like Scottish salmon heading up-river towards their spawning grounds.

This analogy was based on the fact that the only form of orientation recognized in human spermatozoa was rheotaxis, in which the male gametes swim against the current in a moving body of medium. Because the flow of fluid in the female reproductive tract is outward, the spermatozoa inevitably move upstream, in the direction of the ampullary region of the fallopian tubes, where fertilization occurs. Initial contact between the male and female gametes was thought to be the consequence of a chance collision, followed by a highly specific, cell-recognition event that resulted in the tight adhesion of the spermatozoa to an acellular membrane surrounding the egg, known as the zona pellucida.

Ralt and colleagues now propose that the fluid that bathes the oocyte during its dif-



Spermatozoa-egg interaction during mammalian fertilization. The results from Ralt *et al.*, discussed here, imply that the initial contact between human gametes is the result of sperm being attracted by chemotactic factors.

ferentiation within the ovarian follicle (follicular fluid) contains chemotactic factors for human spermatozoa. Their experiments involved the use of two chambers separated by an 8- μ m polycarbonate filter. Spermatozoa were placed in the lower compartment and their ability to migrate through the filters and accumulate in samples of diluted follicular fluid in the upper compartment was assessed. The controls consisted of parallel incubations in which the upper wells contained either culture medium or blood serum, or incubations in which follicular fluid was added to both the upper and lower chambers.

The authors observed a clear accumulation of spermatozoa in the upper chambers in the presence of follicular fluid diluted 1:5,000 to 1:10,000, but no such accumulation in the controls or in dilutions of follicular fluids outside this range. That the accumulation of spermatozoa in the presence of diluted follicular fluid was due to chemotaxis and not a number of alternative explanations was confirmed by looking at the trajectories of human spermatozoa following the micro-injection of follicular fluid into the sperm suspensions. Some of the spermatozoa changed their swimming direction towards the site of the fluid injection. Exactly the same behaviour was observed when egg-conditioned culture medium was microinjected into suspensions of human spermatozoa.

One of the intriguing features of the experiments is the considerable variability exhibited by follicular fluid specimens in their ability to influence the swimming trajectories of the spermatozoa. Indeed, 50 per cent of the samples possessed no activity whatsoever. Such variability would cast severe doubt on the authors' conclusions but for a remarkable correlation between the expression of chemotactic activity by the follicular fluid and the ability of the egg from the same follicle to undergo fertilization *in vitro*. In a blind analysis of 62 follicular fluids from 40 patients there was a highly significant difference in the level of chemotactic activity in follicular fluid samples associated with eggs that did fertilize *in vitro* compared with those that did not.

If confirmed, these observations will have

very many ramifications. The existence of chemotactic factors that accumulate during preovulatory development would provide a means of monitoring the maturational status of oocytes aspirated from the ovary during therapeutic procedures such as *in vitro* fertilization and gamete intrafallopian transfer. Assessment of the chemotactic activity expressed by follicular fluids would be an effective way of monitoring the ability of different ovulation-induction regimes to generate populations of mature preovulatory follicles, and it would also help in identifying causes of female infertility associated with

a failure to generate the factor concerned. Conversely, the observations of Ralt *et al.* imply that receptor sites for the putative factor must be present on the surface of human spermatozoa, the absence of which could be involved in male infertility. Finally, if the chemotactic factor has a part in mediating the union of human gametes at fertilization, the potential for developing radically new approaches to contraception is considerable.

But, as Ralt *et al.* point out, theirs are but preliminary studies which pose as many questions as they answer. What, for instance, is the chemotactic factor? The major follicular steroid, progesterone, is one possibility, because human spermatozoa respond to its presence by generating calcium transients, indicating the presence of some form of progesterone-receptor mechanism at the sperm surface². Ralt *et al.* showed however that chemotactic activity was not correlated with the concentrations of progesterone or oestrogen in the follicular fluids. Other candidates include peptides structurally related to the neutrophil chemoattractant, *N*-formyl-methionyl-leucyl-phenylalanine: receptors for this peptide on the human sperm surface have been described, and a claim made that it induces chemotactic behaviour in human spermatozoa³.

In addition to the physico-chemical nature of the putative chemotactic factor, the stage of oocyte maturation at which it is elaborated, the ontogeny, distribution and structure of the complementary sperm receptor sites and the intracellular mechanisms responsible for the change in flagellar beat characteristics responsible for chemotaxis, all require investigation. Despite the tentative nature of the new work, the biological and clinical implications are such that gamete biologists are unlikely to get much rest in the coming months. □

John Aitken is in the MRC Reproductive Biology Unit, Centre for Reproductive Biology, 37 Chalmers Street, Edinburgh EH3 9EW, UK.

1. Ralt, D. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **88**, 2840-2844 (1991).
2. Osman, R. A., Andria, M. L. & Jones, A. D. *Biochem. biophys. Res. Commun.* **160**, 828-833 (1989).
3. Gnnessi, L. *et al.* *J. clin. Endocrin. Metab.* **63**, 841-846 (1986).

Washing the wind

Air pollution is very hard to control. Once something has escaped into the air, you can't get it back. And yet the various layers of the atmosphere mix only slowly. A lot of smog and pollutant gases linger near ground level for quite a while to plague us all. Some purifying device which made contact with a lot of ground-level air could greatly improve our environment.

The ideal device, says Daedalus, is that epitome of greenery, the windmill. To maximize its efficiency, the blades of a windmill must make contact with as much of the wind as possible. If the blades were coated with a catalyst that oxidized the smog or pollutant on contact, they would purify the air as it passed through.

At first Daedalus thought of adapting the catalysts used to clean up car exhausts. These, however, only work near red heat — a worrying complication in a windmill. Photochemical catalysts based on titanium dioxide seem a better option. In the presence of light and moisture at room temperature, this material converts hydroxide ions to hydroxyl radicals — which oxidize organics rapidly and ferociously. A windmill coated with a hygroscopic titanium dioxide photocatalyst would only work in the daytime. But it would then convert the thickest smog to country-fresh air, and waft it healthily downwind. Under the right conditions, a well sited windmill farm could deliver both power and clean air to the fortunate suburb sheltering in its lee.

Even better, says Daedalus, the oxidation of the air pollutants will help to drive the windmill. Most pollutants, such as carbon monoxide, methane and other hydrocarbons and solvent vapours, liberate a lot of heat when they are oxidized. The product molecules will spring off the catalytic surface with greatly increased thermal velocity, imparting significant thrust to the surface. If the windmill blade is selectively coated with catalyst on its trailing edge, this thrust will help to drive the blade.

To test this elegant notion, DREADCO engineers are painting catalytic surfaces on fans which have to pump very polluted air anyway, like the exhaust fans of road-tunnels, paint shops and chemical works. They should deliver cleaner air for less power — indeed in really dirty air they may even run spontaneously. Once the concept has been proven, Daedalus will offer his catalytic coating technology to the designers of windmills and industrial fans. Even cars, those arch-polluters of our environment, could make amends with catalytically coated cooling fans, while catalytic propeller-driven aircraft could make a token contribution to purifying the upper air. Indeed, says Daedalus, a catalytic coating on their under wing surfaces should not only clean the air they fly on, but slightly increase their lift.

David Jones

Designing better solar cookers

SIR — In their Commentary on solar cooking¹, Kammen and Lankford gave pride of place to the box-type cooker. But this design suffers from several drawbacks, not least that such cookers require cooking to be done in direct sunlight and during the middle of the day, and that they lack storage facilities and have no provision for frying.

In a bid to overcome some of these problems, a cooker with fresnel lens was designed². A large fresnel lens is used to heat a container surrounded by an annular cavity filled with ammoniated salts of magnesium chloride and calcium chloride. Heat is stored chemically in these compounds and is released on demand at 300 °C.

Another system that has storage but is less expensive than the above, uses heat pipes for parabolic as well as flat-plate collectors³. In this arrangement, an evaporator with minimum shading effect is placed at the focus of a solar collector. The energy reaching the evaporator is conveyed rapidly to the condenser end of the heat pipe. The condenser, in the form of a cooking chamber, is located in the shade or inside the kitchen. As the shaded cooking chamber is well insulated, pot losses associated with wind will not arise. This system uses energy from sunshine and will not allow reverse circulation losses, due to its diode-like operation⁴. Because of the unique feature of isothermal operation of heat pipes, temperatures

on a flat-plate collector equipped with a heat pipe will be forced to follow the temperatures in the condenser section.

A more recent addition to the solar cooker family is that developed by Mills and his group at the University of Sydney^{5,6}. This cooker can do practically anything a gas or electric stove can, and because it includes a small pressurized water vessel, it can also heat water or power a refrigerator. It can make two large family meals on sunny days and can easily be supplemented with conventional fuel on cloudy days. It has no moving parts, makes use of common materials and can fry, bake, boil or steam food. Its hot plate can be placed indoors, and heat storage allows cooking after sundown.

A. JAGADEESH

*Society of Science for the People,
2/210, Nawabpet, Nellore-524 002
Andhra Pradesh, India*

KAMMEN AND LANKFORD REPLY — Jagadeesh's letter is similar to several that we have received in response to our article¹. It describes technically sophisticated designs for solar cookers that will achieve higher temperatures than our simple box-cooker. The question is whether they will be used or not.

Not only are the designs technically complex and rather expensive, but one of them uses magnesium chloride and calcium chloride battery salts that could be harmful to livestock or children if the batteries were dumped or broken and the salts consumed.

We strongly support the notion that a diversity of solar and other renewable technology systems is necessary to support energy self-sufficient development, but we feel that Jagadeesh has missed the essential point concerning the ovens. We have found that when the end user (nearly always a woman) makes her own oven, her interest in using it is very high. If someone else makes and donates the oven or significant components (such as ammoniated salt batteries, condensers or pressurized water vessels), her interest is substantially less. For this reason we believe that until solar cooking equipment has become commonplace a design that can be made locally is essential for widespread acceptance of this technology⁷.

DANIEL M. KAMMEN

*Division of Biology, 216-76,
California Institute of Technology,
Pasadena, California 91125, USA*

WILLIAM F. LANKFORD

*Department of Physics,
George Mason University,
Fairfax, Virginia 22030, USA*

1. Kammen, D. M. & Lankford, W. F. *Nature* **348**, 385 (1990).
2. Hall, C. A. *et al.* in *Proc. ISES Congress* (eds F. de Winter & M. Cox) 547 (Pergamon, New Delhi, 1980).
3. Kalifa, A. M. A. *et al.* *Appl. Energy* **24**, 78 (1986).
4. Akyurt, M. *Sol. Energy* **32**, 625–631 (1984).
5. Mills, D. R. *et al.* *Optics Lett.* **5**, 558 (1980).
6. Mills, D. R. & Mao, Y. Q. *Sunworld* **11**, 44–49 (1987).
7. Kammen, D. M. *Natn. geogr. Res. Explor.* **7**, 3–5 (1991).

Olbers' paradox

SIR — At least in a mathematical sense there is yet another solution to Olbers' paradox¹: fractals². Fractal structures like Menger's sponge can have infinite surface area but zero volume and show a self-similar, scaling-independent distribution of their points. Suppose galaxies, clusters and higher order superclusters of galaxies (or, more exactly, the total mass energy including dark matter) would be distributed in such or a similar fractal fashion. Due to the fractal structure bigger volumes have to contain a relatively lower amount of matter and energy. The energy (and thus also radiation) content per volume falls below any given positive boundary only if a sufficiently large total volume is looked at. Even if it should be static, unlimited or eternal, the sky in such a fractal universe remains dark at night. And so, too, does the sky in our own expanding, finite-age Universe, but what about its large-scale structure³; are there some underlying fractal properties?

THOMAS DANDEKAR

*EMBL, Postfach 102209,
W-69 Heidelberg, Germany*

1. Mandelbrot, B. *The Fractal Geometry of Nature* (Freeman, New York, 1977).
2. Efstathiou, G., Sutherland, W. J. & Maddox, S. J. *Nature* **348**, 705–707 (1990).

Prion infection

SIR — The finding that the chaperonin heat shock protein 60 (hsp60) catalyses the assembly of further hsp60 molecules¹ lends weight to the fascinating possibility² that prion infection might be due to self-perpetuating protein-directed changes in topology. Although, as both Carlson *et al.*² and Weissman³ comment, such replicatory topology alteration is unprecedented as a mechanism of infection, the case of hsp60 illustrates that it is not without precedent as a means of protein topology change. The similarity between the suggested mechanism of prion infection and hsp60 activity is perhaps strong enough to suggest that prion protein could be a rogue chaperonin. Further, the hypothesized topology alteration of hsp60 and prion protein suggests that that which replicates need not be nucleic acid.

LAURENCE D. HURST

*Department of Zoology,
University of Oxford,
South Parks Road,
Oxford OX1 3PS, UK*

DAVID HAIG

*Department of Plant Sciences,
University of Oxford,
South Parks Road,
Oxford OX1 3RA, UK*

1. Cheng, M. Y., Hartl, F.-U. & Horwich, A. L. *Nature* **348**, 455–458 (1990).
2. Carlson, G. A., Hsiao, K., Oesch, B., Westaway, D. & Prusiner, S. B. *Trends Gen.* **7**, 61–65 (1991).
3. Weissman, C. *Nature* **349**, 569–571 (1991).

Neotropical plant diversity

SIR — We believe that the overwhelming emphasis on lowland forests with regard to tropical conservation and deforestation has diverted attention from certain patterns of neotropical plant species diversity (see for example, refs 1 and 2). Specifically, the northern Andes, a relatively small and highly deforested region, may have a flora as rich or richer than that of the much larger, mostly intact Amazon basin. Areas of the world that contain the greatest diversity of species and face the most immediate threat of destruction should be given priority for conservation and study.

In most groups of organisms, diversity increases from high to low latitudes and from high to low elevations. The lowland Amazon rainforest has become a popular symbol of a beleaguered stronghold containing millions of species, most of them unknown to science. But the Amazon basin is far from the richest region in the neotropics when plants are considered. Other, smaller regions, such as the northern Andes, Central America and the Atlantic coastal forest of Brazil, are comparatively as rich or richer, and are almost completely deforested, whereas the Amazon forest is relatively intact (see table).

There is no published account of all neo-

DEFORESTATION AND DIVERSITY IN THE NEOTROPICS				
Phytogeographic region*	Amazon basin	Northern Andes	Atlantic coastal forest of Brazil	Central America and Mexico
Surface area in sq km†	7,050,000	383,000	1,000,000	2,515,295
Area deforested (%)‡	8–11	90–95	>95	cs 60
Estimated total number of species §	30,000	40,000	10,000	19,000
Flora neotropica Sample II	1,347	891	793	1,038
Gentry's sample ¶	1,829	1,454	1,288	1,934

* The four most diverse neotropical phytogeographic regions from Gentry⁷; elevational boundaries between regions are the 500 m contour.

† Surface area is plane surface and does not take into account topography.

‡ Figures from: Amazon basin (Brazil only)⁸; Northern Andes^{9,10}; Atlantic coastal forest of Brazil¹¹; Central America (excluding Mexico) from G. Hartshorn, personal communication.

§ Figures from Gentry^{7,9} and Myers¹¹. Figure for Central America excludes Mexico.

¶ Sample taken from 43 volumes of *Flora Neotropica* (flowering plants only) dealing with 5,331 species (that is about 6 per cent of estimated total of 90,000).

¶ Sample taken from Gentry⁷.

tropical plants, and even the total number of species is unknown. Many botanists accept an estimated total of 90,000 species of flowering plants. There are some estimates of how this total may be distributed amongst the various neotropical phytogeographical regions. We have used these, as well as the results of two samples, to get an idea of the relative diversity of each region (see table).

The table shows that, for its size, the northern Andes is the most diverse region in the neotropics, and its forests have been progressively reduced for centuries. Although this region has a surface area only one twentieth that of the Amazon basin, it contains at least as many species of flowering plants. Among the bryophytes, 93 per cent of the approximately 900 species of moss known from Colombia are found in the Andean region, which accounts for only 20–25 per cent of the Colombian land surface; the remaining 7 per cent are restricted to the surrounding lowlands. Only 200–250 species of moss occur in the entire Amazon basin. Hepatics show similar trends in Colombia, approximately 80 per cent of the species being restricted to the Andes³. Among pteridophytes, the Andean region contains approximately 1,500 species of fern and fern allies, whereas, in contrast, the Brazilian Amazon basin contains only 280 species⁴. The homosporous ferns are also particularly numerous in the Andes: approximately 1,000 species are recorded from Ecuador (300 of which are from the eastern lowlands). The figures are smaller for lowland neotropical countries, for example, 260 for Suriname and 220 for Brazilian Amazonia⁵.

Much of the data in support of the diversity of lowland neotropical forest have come from recent inventory studies. Spectacular results have been recorded. For example, 300 species of tree with a diameter at breast

height of at least 10 cm have been found in one hectare in Amazon Peru⁶. We do not question that there are many tree species in lowland neotropical forests, but our argument is for all plants and not just trees. By using the standard sampling technique that includes only woody plants with stem diameters equal to or more than 10 cm (usually in 1 hectare sites), most of the plants present are probably missed. Unfortunately, there is not, to our knowledge, any inventory of primary forest that takes account of all plant species, including bryophytes.

Our purpose here is to draw attention to the diverse, threatened flora of the northern Andes, which we feel attracts disproportionately little interest from biologists and conservationists. We do not understate the importance of the Amazon basin and its biota. But figures such as 8–11 per cent total deforestation in the Amazon basin pale in significance when one considers the northern Andean region.

ANDREW HENDERSON

STEVEN P. CHURCHILL

JAMES L. LUTEYN

*New York Botanical Garden,
Bronx, New York 10458–5126,
USA*

1. Wilson, E. O. (ed.) *Biodiversity* (National Academy Press, Washington, D.C., 1988).
2. Gentry, A. H. (ed.) *Four Neotropical Rainforests* (Yale University Press, New Haven, 1990).
3. Churchill, S. P. *The Bryologist* (in the press).
4. Tryon, R. & Tryon, A. F. *Ferns And Allied Plants With Special Reference To Tropical America*. (Springer, New York, 1982).
5. Tryon, R. *Bot. Rev.* **52**, 117–156 (1986).
6. Gentry, A. H. *Ann. Missouri Bot. Gard.* **75**, 1–34 (1938).
7. Gentry, A. H. *Ann. Missouri Bot. Gard.* **69**, 557–593 (1982).
8. Bonalume, R. *Nature* **345**, 754 (1990).
9. Gentry, A. H. in *Floristic Inventory of Tropical Countries* (New York Botanical Garden, New York, 1989).
10. *Mapa de Bosques de Colombia*. Instituto Geográfico 'Augustin Codazzi', Bogotá (1985).
11. Myers, N. *The Environmentalist* **8**, 1–20 (1988).

Is cold dark matter really dead?

SIR — Our article in *Nature*¹ and the accompanying News and Views² appear to have been widely interpreted as ruling out the existence of cold-dark-matter. Although the article makes it quite clear that it is the standard cold-dark-matter model that is questioned by our data, we are concerned that workers in other fields may not fully appreciate the distinction. The standard model is based on four premises: (1) that dark matter consists of weakly interacting 'cold' elementary particles; (2) that the Universe has critical density; (3) that the seed fluctuations from which cosmological structure grew are of the type predicted by the inflationary theory of the early Universe and (4) that the distribution of galaxies is related to the distribution of mass through a simple statistical prescription, the 'linear biasing model'.

The discrepancy between the size of the superclusters in our map of galaxies detected by the Infrared Astronomical Satellite (IRAS) and the predictions of the standard model imply that one or more of the model assumptions is incorrect. It does not automatically rule out the existence of cold dark matter. Indeed, in two separate studies^{2,3}, we have shown that the dynamics of the superclusters indicates that the Universe has a density very close to the critical value — as hypothesized in (2) above. Standard Big Bang nucleosynthesis limits the density of baryons to less than about 10 per cent of the critical value. Thus, our data from IRAS provide strong evidence that the Universe is dominated by non-baryonic dark matter.

A viable cold-dark-matter theory might grow out of the standard model through a more physical biasing prescription or alternative seed fluctuations, perhaps associated with cosmic strings or textures. Exploring these and other possibilities is one of the exciting challenges facing cosmologists today. Ultimately, however, the cold-dark-matter paradigm will stand or fall on the basis of the laboratory searches, now underway in the United Kingdom and elsewhere, for the elementary particles it postulates. We believe that these searches should continue to be energetically pursued.

CARLOS S. FRENK

*Physics Department,
University of Durham,
South Road,
Durham DH1 3LE, UK*

NICK KAISER

*CIAR Cosmology Program and
Canadian Institute for Theoretical
Astrophysics,
Toronto,
Ontario M5S 1A1, Canada*

1. Saunders, W. et al. *Nature* **349**, 32–38 (1990).
2. Lindley, D. *Nature*, **349**, 14 (1990).
3. Rowan-Robinson M. et al. *Mon. Not. R. astr. Soc.* **247**, 1–19 (1990).
4. Kaiser N. et al. *Mon. Not. R. astr. Soc.* (in the press).

Legislating for clean water

Eric Ashby

The Limits of Law: The Public Regulation of Private Pollution. By Peter C. Yeager. Cambridge University Press: 1991. Pp.369. £35, \$49.50.

In the 1970s, the US Army Corps of Engineers was about to construct a \$1.2 billion hydroelectric dam on the St John River in Maine. The river is not in a region of virgin wilderness. It is an area for potato farms and woodpulp mills. But on its steep banks, among spruce and alder, there grows a rare plant, the Furbish lousewort, named after the amateur botanist who discovered it. It is a modest plant, but under the Federal Endangered Species Act it is protected. Accordingly, in 1977 President Carter stopped further work on the dam until the future of the Furbish lousewort could be assured.

How did the US government, cradle of the military-industrial complex, come to pass a law like that, constraining our exploitation of nature, often much to our inconvenience? In *The Limits of Law*, Peter Yeager examines this question as it applies to laws to control industrial water pollution in the United States. He records in detail the growth of federal law to clean up rivers, culminating in the historic and euphoric *Federal Water Pollution Control Act Amendments* in 1972; an Act historic because it was passed by Congress in the teeth of a veto by President Nixon; euphoric because it aspired to make all rivers in the United States "fishable and swimmable" by 1985.

Yeager's detailed account is attached to a theme, namely that the drafting of anti-pollution laws, the regulations made under them, and the commitment to enforce them, were all limited by the bias of politicians, and self-interest of polluters, and latterly by the greening of public opinion. A subsidiary theme is that these limits were imposed under the influence of élites, especially large corporations, often at the expense of small business and the general public. As background to his essay, Yeager discusses how laws are made to regulate social activities like industry. He concludes that they are not the product of consensus between lawmakers and those they seek to regulate; they are the product of conflict between industrialists who want to exploit the environment and politicians who want to safeguard the interests of other voters as well as to keep the good will of industrialists. Participating on both sides of the conflict are technical experts, whose advice may set limits to the whole process, from drafting of the law to enforcement of the regulations. Supported by anecdotal evidence and some dubious regression coef-

ficients, Yeager gives the impression that big industry has had more than its fair share in setting limits to cushion the effects of legislation, to make the wording of some laws 'symbolic', the regulations ambivalent, and the enforcement supine.

But the constraints that limit law making and its enforcement do themselves change. Thus, for a time industrialists called the tune, but when toxic chemicals got into the drinking water, consumers joined the conflict and

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Discharge pipeline from a large pharmaceutical company.

influenced the limits more in their favour. The same law may then be reinterpreted by the officers in the field who have to enforce it. At this stage industrialists develop what Yeager quotes as 'corporate responsibility', taking a lead in pollution control to restore the confidence of the public.

All this is familiar to anyone concerned with environmental policy in Europe and Yeager does not make any fresh contribution to it. The interesting part of his book is the detailed account of the debates leading to the critical law passed in 1972 and its aftermath. Such an enterprising law needed an enterprising bureaucracy to administer it. That was to hand. The Environmental Protection Agency (EPA) had been set up in 1970. It had 6,000 employees and an initial budget of \$1.4 billion. The first task of its Water Quality Office was to get data on pollution from thousands of industrial polluters and to decide how much each of them should be allowed to discharge. Industrialists then challenged the proposed consents. By 1973 there were still over 1,100 requests for hearings pending. In its initial enthusiasm the recommended to the Justice Department that lawsuits should be launched against 200 corporations. Meanwhile, there was confusion about the way controls should be achieved.

The original recommendation was for the use of the Best Practicable Technology (BPT). But of course the word 'practicable' is a bonus for lawyers in court and there was pressure, when it was disclosed that carcinogens were among the pollutants, to abandon BPT for BAT — the Best Available Technology.

In the first flush of its activities the ugly words 'cost-benefit' were not audible in the corridors of EPA. But lobbying in Washington on the part of corporations alerted the Office of Management and Budget (OMB) to the hazards to industry of an environmental agency that was too zealous. It was suspected that OMB "often stalled the EPA's legislative proposals and testimony to Congress, giving advantages to industry . . ." and "might even be tampering with technical data" which supported EPA's case for stricter control. It was President Reagan who finally sabotaged EPA's strategy. Private sector costs were to take precedence over public-sector benefits. EPA's budget was cut. Its administrator was told to submit cost-benefit statements for all regulations. This amounted to deregulation. That was in 1981 and arguably it has handicapped American environmental policy ever since. In the end it has been the courts of law which have defended EPA against the erosion of its powers. But deregulation, too, has its limits and constraints. In the 1980s there was a massive legal action brought by a group of major corporations against the

EPA, challenging its estimates of the cost of pollution control. They thought they were riding on the wave of deregulation. They were wrong: in April, 1989 the appellate court unanimously rejected their challenge, and upheld the EPA rules.

It is a reviewer's duty to issue a warning about *The Limits of Law*. Readers of books like this do not expect them to be written in polished prose. They are prepared to read, two or three times over, sentences conveying subtle and elusive ideas. But readers do expect straight reportage (which is the main material in this book) to be intelligible at first reading. Peter Yeager does not meet this expectation. His sentences sag under the sheer weight of abstract words. There are rich examples of what H. W. Fowler in *Modern English Usage* (Clarendon, 1965) called "abstractitis". It is a style of unparalleled viscosity. The effort of reading it can be compared to the effort of walking through a swamp in over-sized gum boots. It is a pity that the book is so exhausting to read, for Yeager has an interesting and important story to tell. □

Lord Eric Ashby is at 22 Eltisley Avenue, Cambridge CB3 9JG, UK.

In the eye of the beholder

J. Anthony Movshon

Vision: Coding and Efficiency. Edited by Colin Blakemore. Cambridge University Press: 1991. Pp.448. £65, \$120.

IN a scientific career spanning five decades, H. B. Barlow — one of the seminal figures in modern sensory physiology — has contributed extensively both to our understanding of visual processing and to the way in which we conceptualize the brain's representations of sensory information. *Vision: Coding and Efficiency* — a tribute to Barlow — derives from a symposium held in Cambridge in 1987 to celebrate his 65th birthday and his concurrent retirement from his Royal Society professorship, if not his retirement from science (he remains active today). Edited by Colin Blakemore, Barlow's first student, this handsomely produced book spans the extraordinary range of fields in which Barlow has made his presence felt.

The range of topics represented is both the strength and the weakness of the undertaking. Unlike many *festschriften* for lesser figures, which tend to collect too many papers devoted to a narrow subfield, this effort has a broad coverage of many areas of modern visual science, from fundamental optics, psychophysics and neurophysiology through conceptual models of image understanding. The result is almost the skeleton of a textbook of visual science. The only difficulty is that the coverage is so broadly based on the interests of those who have been associated with Barlow over the years that it is inevitably a bit uneven.

Nonetheless, there is much to treasure. The roster of authors is a distinguished one indeed, and a surprisingly large proportion have produced papers that are not simply minor revisions of their standard 'canned'

review article. Notable among these are D. G. Pelli's lead chapter on quantum efficiency, which lays out an elegant framework (owing much to Barlow) for understanding the concept of statistical efficiency as applied to vision. D. Kersten's chapter is a brave effort to extend this kind of concept to higher-level processes of image understanding. Other papers on various aspects of spatial vision, light and dark adaptation, and colour vision, all add to a satisfying and coherent group of readings on the analysis of visual function by psychophysical techniques.

Coverage of visual neurobiology and development is a bit patchier. The book contains an intriguing mixture of papers on comparative topics (noteworthy and elegant are contributions by J. D. Pettigrew on stereopsis and by M. F. Land on compound eyes), on retinal function and organization (among them an appealing but stylistically dissonant piece on retinal heterogeneity by D. I. Vaney and A. A. Hughes), and on cortical processing (among them useful developmental papers by C. Blakemore and by R. C. Van Sluyters and his colleagues, and an extensive review of recent work on cortical architecture by S. M. Zeki). Because this part of the collection seeks to cover a vast and growing field, however, it leaves open rather more gaps than it fills.

Many of Barlow's most enduring contributions have been theoretical, and theory has its place here too. Again one might quibble at some unevenness and eccentricity; A. B. Watson's alarmingly crystalline concept of visual cortical processing stands out as a provocative effort in a slightly mundane set of chapters, as does Barlow's own contribution, a characteristically thought-provoking piece on the basis for visual after-effects.

It is a little hard to decide for whom this book is best suited. It has many strong chapters, especially on psychophysical topics, and these form a state-of-the-art coverage of an important part of the field. In other areas, things are more scattered, so that the book as a whole cannot be endorsed for an audience seeking a full general treatment. But that would be a great deal to ask of any volume of this sort. Perhaps the ideal reader of the book has the intellectual breadth and curiosity of Barlow himself, and the same inclination to absorb concepts and ideas from a loosely-knit but often fascinating collection of disciplines. □

J. Anthony Movshon is at the Center for Neural Science, New York University, 6 Washington Place, 8th Floor, New York, New York 10003, USA.

The bare bones

S. Mann

Skeletal Biomineralization: Patterns, Processes and Evolutionary Trends. Volumes I and II. Edited by J. G. Carter. Van Nostrand Reinhold: 1991. Pp.1032. \$99.95, £79.

If you were choosing a book to slip into your briefcase intending to pull it out again during a long international flight to an annual conference, it would not be *Skeletal Biomineralization, Processes and Evolutionary Trends* edited by Joseph Carter. For a start, at over 1,000 pages, the book would make a large hole in your baggage allowance. More importantly, reading a book like this is a serious matter.

The book is big in every way, and I have to say that I find the editor's dedication to producing this tome as awe-inspiring as the subject matter itself. The book comes in two volumes. The first 600 pages of volume I contains 25 invited chapters overviewing the mineralogy and microstructure of calcified skeletal tissues. The scope of these chapters vary enormously; the first nine contributions are relatively short (8–15 pages) and cover various aspects of biomineralization including general mechanisms, biomechanics, and calcification in spicules, corals, sponges, arthropods and gastropods, for example. Chapters 10–16 are written by the editor and are concerned with shell microstructure — over 270 pages of immense detail. There-

after, several shorter overviews, on subjects such as protochordate biomineralization, vertebrate skeletal tissues, dental structures and geology and related applications of isotopes, trace elements and amino acids, widen the perspective.

But all is not finished; another 200 pages including a glossary and a bibliography come before the index. The bibliography must rank as the best ever collated in this field. It is over 100 pages long and includes its own index. Volume II is more modest in size and provides supplementary illustrations, mainly in the form of photographs, of the skeletal structures discussed in volume I. The quality of production of electron micrographs, like that in volume I, is first class.

I was very impressed by this book. The work is scholarly and it is a mark of real scientific endeavour. Although the subject matter is very specialized, being at the more geological end of biomineralization, I recommend that chemists, materials scientists and biologists look at this book. The electron micrographs of volume II testify to the sophistication of biological processes to synthesize and assemble inorganic materials that would be termed smart in other more fashionable disciplines. This is a vast unexplored area and this book will be a valuable contribution not only in the detailed classification of skeletal structures but as a source of wonder and inspiration to scientists interested in the possibilities of mimicking biological materials in the laboratory. □

S. Mann is at the School of Chemistry, University of Bath, Bath BA2 7AY, UK.

Pulitzer Prize

This year's Pulitzer Prize for general nonfiction has been won by zoologists Bert Hölldobler of the University of Würzburg and Edward O. Wilson of Harvard University for their monumental work *The Ants*.

The book, which was published in 1990 by Harvard University Press and Springer, runs to an astonishing 745 pages, has over 1,000 figures and includes all aspects of ant biology, evolution and ecology, with an emphasis on the reasons underlying the success and importance of the group. According to Donald Feener, Diane Davidson and Jon Seger who reviewed the book in these pages (see *Nature* **344**, 894–895; 1990), "Only Hölldobler and Wilson could have written such a comprehensive and integrated treatment of ant biology. It represents a herculean labour of love, and it sets a new standard for synthetic works on major taxa."

Flights of the season

Jeremy J. D. Greenwood

Bird Migration. By Thomas Alerstam. Cambridge University Press: 1991. Pp.420. £55, \$105.

MORE than anything else, the ability to fly characterizes birds. It imposes on them severe anatomical, physiological and ecological constraints. But it also gives them freedom — not just in a physical sense but ecologically, for they are able to exploit dispersed, patchy and variable resources in a way quite impossible for us earthbound mammals. Starlings clear one bird table and move on to the next; in the countryside they move from field to field, sometimes completely changing the direction in which they fly out from their roosts on successive days, as birds that encountered diminishing food supplies the day before follow their more successful fellows. In winters when their food supplies fail, the seed-eating birds of the northern forests, redpolls, crossbills, nutcrackers and waxwings, move far outside their normal ranges in search of food. Ptarmigan, breeding high in the mountains in summer, move down the slopes as winter advances. In Africa, queleas follow behind the moving rainbelts, so that they can feed on the flush of seeds produced in the few months after rain has fallen.

Virtually all bird migrations are seasonal. As Thomas Alerstam puts it in *Bird Migration*, "Birds migrate because we are tilted at 23.5°." He explains lucidly how this tilt and other aspects of the Earth's yearly round produce both the seasonal and geographical changes in temperature and rainfall that are at the root of migrations and the great wind belts that are so important in determining how particular journeys are made.

The journeys are often stupendous. Arctic warblers from southeast Asia, weighing under 10 grams, move 9,000 kilometres north to breed in the short summer of northern Scandinavia, much of their flight over the scarcely welcoming terrain of central Asia. Each autumn, thousands of millions of birds flood into subsaharan Africa from Europe and Asia. Typically, those from Europe cross the Mediterranean, North Africa and the Sahara in one hop of 50–70 flying hours. North Africa is too dry in autumn for it to be worthwhile stopping and the Saharan oases too few and too tiny. There is even evidence that sedge warblers fly directly to West Africa from southern England, a 70–90-hour journey at 50–60 kilometres per hour. They are generally boosted by favourable tail-winds but even so need to lay down so much fat and extra muscle that their weight at the start of the journey may be almost double its normal value. The winds are largely unchanged, and therefore adverse, in

the spring but North Africa is now verdant, allowing the journey to be made in two stages. Similar journeys, but entirely over water, occur in the Americas. Columbus encountered large flocks of land birds midway between Bermuda and Puerto Rico, part of the great migration across the west Atlantic, which includes some small birds flying non-stop the 4,300 kilometres from Nova Scotia to Venezuela.

The exodus of many species, especially insectivores, from arctic, boreal and even temperate regions at the end of the summer is easily — perhaps too easily — explicable in terms of deteriorating food supplies and physical conditions. But why should they ever move north in the first place? If Africa can support them for part of the year, why not for the whole year? Does the answer simply lie in the otherwise unexploited abundant insect populations of the northern summer?

Turning to the journeys themselves, Alerstam gives a clear exposition of the ways in which birds fly and of how these influence migration strategies. The absence of thermals over water presents us with the great migration spectacles of thousands of soaring birds concentrating at narrow sea crossings at Gibraltar, the Bosphorus and elsewhere. He explains why birds use fat as fuel and how their varying abilities to store and use it determine the length and form of their journeys. He gives a detailed and fascinating account of how birds cope with vagaries of wind and weather, not just choosing under what conditions to migrate but responding during their flight to changing conditions and adaptively varying the altitude at which they fly. But even birds are not perfect and many are driven off course by adverse winds, to delight rarity-hunting birdwatchers.

There are other hazards too. Migrants

Ardea

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Come fly away — flocks of waders setting out from Kent.

Or is competition with African residents important? (And is competition with the migrants a reason for the low densities of some African residents?) Alerstam touches on these ecological questions largely on a case-by-case basis, in a long section in which he considers the migrations of individual species, grouped according to habitat and food. The details and diversity are fascinating, the anecdotes inspiring. Perhaps the diversity is too great for many broad principles to be extracted but one is left longing for a more synthetic treatment of the ecological factors that shape migration. Alerstam, not unreasonably, concentrates on the birds breeding in northwest Europe that he knows so well but one wonders if more ecological insights might emerge if one was to think of them not as northern birds that winter in the south but as southern species that spend a brief period of the year breeding in the north.

crossing the Mediterranean have to run the gauntlet of Eleonora's Falcons, their own breeding delayed so that they can exploit the full potential of the huge autumn food supply. And some just get lost. Alerstam gives a fairly brief (though comprehensive and balanced) account of how birds orientate and navigate. The relatively low emphasis on this aspect is fortunate, for this is a fast-moving area of research and the book has been translated with minimal updating from the Swedish original, published in 1982. That it is somewhat out of date is a disadvantage but it remains useful. It presents both the basic principles and the fascinating natural history — and it does so in a style that preserves the wonder and magic of its subject. □

Jeremy J. D. Greenwood is at the British Trust for Ornithology, The Nunnery, Nunnery Place, Thetford Norfolk, IP24 2PU, UK.

The Earth's changing face

Mike Kirkby

World Geomorphology. By E. M. Bridges. Cambridge University Press: 1990. Pp. 260. Hbk £35, \$70; pbk £13.95, \$24.95.

Temperate Palaeohydrology: Fluvial Processes in the Temperate Zone during the last 15,000 Years. Edited by L. Starkel, K. J. Gregory and J. B. Thornes. Wiley: 1991. Pp. 548. £70, \$175.

As E. M. Bridges says in his introduction to *World Geomorphology*, there have been sweeping changes in understanding the Earth's crust since 1960 and some reluctance on the part of geomorphologists to digest the implications for the Earth's surface skin. Here was a perfect opportunity to provide new interpretations of major landform units, taking into account current views of plate tectonics, the great activity of the crust and our rapidly changing Pleistocene climates.

At continental scales, large-scale plate movements associated with sea-floor spreading provide passively extending continental margins like those around the Atlantic, and actively compressing margins as around the Pacific. Along passive margins, erosion and deposition lead to continental unloading and submarine loading, with consequent isostatic uplift and subsidence constrained by flexing of the elastic crust. These processes lead to the characteristic uplift of landward-dipping escarpments, increasing seaward dips in the rocks of continental coastal plains, and to the containment of catchments like the Mississippi or the Amazon which drain a large part of continental interiors. If thermal conditions in the crust are suitable it may eventually lead to progressive growth of the continent with the start of an active subduction zone.

At active margins, volcanic activity and compressive folds built mountain belts which are eroded to provide sediment which may be repeatedly recycled as it is returned to the continental margin on the subducting ocean plate. As the mountains are uplifted, their rocks are weakened by severe fracturing and the high elevations are also responsible for exceptionally high erosion rates, due to greatly increased precipitation and perhaps glaciation. Eventually the mountains become so steep and high that erosion rates balance uplift rates, perhaps even at the 20 mm per year reported for the New Zealand Alps. At smaller scales, the key interactions are between local tectonic stress fields, fault patterns and erosion, and there is much to be done before a clear understanding emerges, although some patterns may be seen, for example in block-faulted or rifted areas.

How far has Bridges achieved a worthwhile integration at the broad-scale implied

by his title? Instead of a dynamic account, perhaps concentrating on those areas where most is known, we have an introductory chapter on plate tectonics at a broadly descriptive level, followed by a chapter on each of the six continents and on the Pacific Basin. For each continent, there is a token introduction on plate tectonics but rarely any dynamic view of the structure and form of the land. In fact there is very little fresh analysis and much that could have been written 50 years ago. The basic regionalization of the United States, for example, where much is now known, is still based on N. M. Fennerman's analysis of the 1930s and there is nothing on the dynamic behaviour of the mountain masses of the California coast.

Of course there are difficulties in attempting a region-by-region account of landforms across the world, but the resulting book frequently declines to the level of a catalogue of areas, with no space to give more than a bare description of each, and certainly no chance to present or discuss new insights. I hope that another author will take up the challenge.

Temperate Palaeohydrology edited by L. Starkel, K. J. Gregory and J. B. Thornes might appear to have a similar sweep in scope, with its 15,000-year time span, but its aims are more modest in scope. It is the concluding volume in a series of edited contributions arising from a project undertaken by the International Geological Correlation Programme over the last decade. The theme is introduced by chapters on river hydrology and the sedimentology of river deposits. Over half of the book consists of a series of regionally based case studies which span

Europe. There are three studies on Britain, two each on Finland, Benelux, Poland and the Danube basin, and one each on the eastern Baltic, Switzerland and western Siberia. In all cases the crucial evidence relates to dating and correlation of buried channels and terraces with other forms of evidence for changes in relative sea levels and erosion rates. Several of the studies are for areas of glacial isostatic rebound. Another group is for areas influenced by glacial meltwater or ice damming, and there is a group of rivers which were not directly influenced by glaciation.

The final 150 pages of the book provides some broader perspectives. The most substantial is a survey of methods for retrodicting former discharges from the sedimentological record, with the implications for the associated hydrological balances for the immediately postglacial period, when discharges were several times greater than today. The changes are attributed mainly to the lower evapotranspiration in cooler conditions rather than to major changes in precipitation. There are also useful perspectives on glacier fluctuations, and on the possibilities for broad correlations across the whole area of Europe, among others. Like many other contributed collections, *Temperate Palaeohydrology* gives a fair view of the state of the art in a particular field, but emphasizes the diversity of activity rather than providing broad insights which could be carried back to the global scale attempted in *World Geomorphology*. □

Mike Kirkby is in the School of Geography, University of Leeds, Leeds LS2 9JT, UK.



Pacific paradise — Dutchman Jacob Roggeveen discovered Easter Island in 1721 when the famous stone figures still had the headpieces intact. The island's people apparently believed their nearest neighbours were on the Moon. Taken from *The Royal Geographical Society History of World Exploration* edited by John Keay (Hamlyn, \$20, \$29.98).

Thermodynamic regulation of ocean warming by cirrus clouds deduced from observations of the 1987 El Niño

V. Ramanathan & W. Collins

Scripps Institution of Oceanography and California Space Institute, University of California at San Diego, La Jolla, California 92093, USA

Observations made during the 1987 El Niño show that in the upper range of sea surface temperatures, the greenhouse effect increases with surface temperature at a rate which exceeds the rate at which radiation is being emitted from the surface. In response to this 'super greenhouse effect', highly reflective cirrus clouds are produced which act like a thermostat, shielding the ocean from solar radiation. The regulatory effect of these cirrus clouds may limit sea surface temperatures to less than 305 K.

WATER vapour and clouds are the dominant regulators of the radiative heating of the planet. General circulation model studies have identified the complex effects of the various cloud and water vapour feedbacks and their importance to climate change¹⁻³. We do not understand how these radiative effects respond to a climate change, nor do we understand their feedback effects. What we need is a natural climate-change experiment in a region that exhibits a strong coupling between surface temperatures and radiative fluxes. The tropical Pacific shows such behaviour. In particular, during the El Niño events which occur once every 2-6 years⁴, the equatorial Pacific warms by as much as 2-4 K, and the warming is accompanied by marked variations in the radiation fluxes at the top of the atmosphere⁵. Accurate data on the radiation budget from the Earth radiation budget experiment (ERBE)⁶ and sea-surface-temperature (SST) data from weather satellites and ships⁷ are available for the 1987 El Niño⁸. We use this data set to explore the link between ocean warming and the radiative feedbacks. Relevant definitions are given next.

The total greenhouse effect of the atmosphere and clouds, G , is defined⁹ as $G = E - F$, where F is the radiation emitted to space, and $E = \sigma T^4$ is the energy emitted by the ocean surface, with $\sigma = 5.67 \times 10^{-8} \text{ W m}^{-2} \text{ K}^{-4}$ and $T = \text{SST}$. Observations of F over the clear-sky regions, F_c , are employed to obtain the atmospheric portion of the greenhouse effect, $G_a = E - F_c$. The difference between G and G_a gives the enhancement of the greenhouse effect caused by clouds, or 'cloud long-wave forcing', $C_l = G - G_a = F_c - F$. Clouds also enhance the albedo (reflectivity) of the planet, thereby decreasing the planetary solar heating. This effect, the 'cloud short-wave forcing' (ref. 6), C_s , is estimated by subtracting the solar energy absorbed by a region from that over just the clear-sky portion of that region. Hence, the absorbed solar energy can be expanded as $S(1 - A) = S_c + C_s$, where S is the solar radiation at the top of the atmosphere, A is the column albedo, and S_c is the clear-sky solar absorption. ERBE measurements yield the fluxes for the top of the atmosphere of F , F_c , S and S_c together with the albedo A . Thus we have observable definitions for the various feedback terms.

During the El Niño of 1987, the equatorial Pacific warmed by as much as 3 K (Fig. 1a). The warming was accompanied by a significant increase in the greenhouse effect of the atmo-

sphere (Fig. 1b) and that of the clouds (Fig. 1d). These positive feedback effects are offset by a significant decrease in the absorbed solar energy (Fig. 1c). We show that this cirrus regulation of ocean warming is triggered primarily when SSTs exceed 300 K. Results of past studies⁹⁻¹³ enable us to conclude that the interactions between SST, deep convection, large-scale transport of moisture and cirrus formation are responsible for the observed negative feedback effect of cirrus.

Analyses of the interactions

We begin with the net radiative heating of the surface-atmosphere system, $H = S(1 - A) - F$. Adopting the previous definitions for $F = \sigma T^4 - G_a - C_l$ and $S(1 - A) = S_c + C_s$, we have

$$H = S_c - \sigma T^4 + G_a + C_l + C_s \quad (1)$$

The absorbed solar radiation, S_c , heats the surface, which emits the energy as long-wave radiation, σT^4 . An amount G_a is

TABLE 1 Summary of feedback terms

a, Experiment 1: spatial and seasonal				
Domain	Period (month/year)	dG_s/dT ($\text{W m}^{-2} \text{K}^{-1}$)	dC_s/dC_l	
25° N-25° S	4/85	5.6†	-1.07*	
	4/87	6.5†	-1.03†	
	2/87	6.9†	-0.96*	
	2/88	6.3†	-0.94*	
10° N-10° S	4/85	7.2†	-1.13*	
	4/87	8.1*	-1.12*	
	2/87	8.2*	-0.9†	
	2/88	8.5†	-0.9†	
0-25° N	2/85	4.6*	-1.01*	
	4/85	5.2†	-1.26*	
0-25° S	2/85	7.4†	-1.03†	
	4/85	6.5†	-0.87*	
10° N-10° S	4/85; 2/87; 2/88; 4/89	7.7*	-0.95†	
Atlantic	Same	7.9*	-1.03†	
Indian Ocean	Same	8.1‡	-1.03†	

b, Experiment 2: El Niño						
Domain	Period (month/year)	dG_s/dT ($\text{W m}^{-2} \text{K}^{-1}$)	dC_s/dC_l	dC_s/dG	a_s ($\text{W m}^{-2} \text{K}^{-1}$)	a_l
10° N-10° S	{4/87-4/85; 4/89-4/87; 2/88-2/87}	6.8*	-1.2*	-0.92†	-22‡	18*
	4/87-4/85	9.2*	-1.33†	-0.97*	-20*	16*
	4/89-4/87	8.5*	-1.23†	-0.95†	-19†	17*
	2/87-2/85	6.4‡	-1.09†	-0.97†		
	2/88-2/87		-1.11*	-0.97*		
	5/87-5/85	8.0*	-1.22†	-0.92†		
	7/87-7/85	5.7*	-1.17†	-0.79‡		
	10/87-10/85	6.0*	-1.32†	-0.99*		

Unless otherwise stated, the domain is the Pacific Ocean stretching from 140°E (western Pacific) to 90°W (eastern Pacific). SST is denoted by T ; the atmospheric greenhouse effect by G_a ; the cloud long-wave forcing is C_l ; the total greenhouse effect is $G = G_a + C_l$; the cloud modulation of absorbed solar energy is the cloud short-wave forcing C_s ; and the cloud-forcing slopes a_s (short-wave) and a_l (long-wave) are obtained from Fig. 4a and b, respectively.

* Correlation coefficient $|r|$ for least-squares linear fit is in range $0.9 > |r| > 0.8$

† $|r| > 0.9$

‡ $0.8 > |r| > 0.7$. When $|r| < 0.7$, the correlation is considered insignificant and is omitted from the table.

trapped in the atmosphere, thus reducing the cooling to space. The long-wave energy flow out of the system is further reduced by the flux C_1 trapped in clouds. But this heating effect is offset by the reduction C_s of solar absorption by clouds with $C_s < 0$. We wish to examine how G_a , C_1 and C_s vary with T . These dependences are obtained both from the spatial and seasonal variations in T and from the 1987 El Niño, which began during late 1986.

We use ERBE measurements for 1985–89. The SSTs are obtained from combined analyses from the National Oceanic and Atmospheric Administration and the National Meteorological Center⁷ of satellite and *in situ* data. The data are monthly averaged values with a spatial resolution of 2.5° (latitude) $\times$ 2.5° (longitude). The analyses will focus primarily on April, but we will briefly review results for the months of February, May, July and October. April is preferred for several reasons. Because of the near-hemispherical symmetry of solar radiation and of meteorological features in the equatorial regions, we can safely ignore hemispherical asymmetries in the dependence on the solar zenith angle before comparing the albedos of the different oceans. Around April, equatorial SSTs reach peak seasonal values and exceed 300 K over most of the Pacific⁴, the minimum SST necessary to initiate deep convection. Finally, the 1987 El Niño warming stabilized at near-maximum values around April 1987.

Experiment 1

In this experiment, we examined spatial variations in the greenhouse effect and cloud forcing for different seasons. The total

greenhouse effect $G = G_a + C_1$ increases sharply when the SST exceeds ~ 300 K (Fig. 2a), a feature that is present in all the months and can also be inferred from earlier studies^{14,15}. Optically thick cirrus clouds in the upper troposphere are needed to account for the large increase in G of $\sim 100 \text{ W m}^{-2}$ as the temperature increases from 300 to 303 K. The slope, dG_a/dT , for the tropical Pacific (10° N to 10° S) is $\sim 6\text{--}9 \text{ W m}^{-2} \text{ K}^{-1}$ (Fig. 2b, and Table 1), whereas it is $\sim 3.4 \text{ W m}^{-2} \text{ K}^{-1}$ for the entire Pacific (60° N to 60° S). About half of the $3.4 \text{ W m}^{-2} \text{ K}^{-1}$ is due to atmospheric trapping of the increased black-body emission. The balance is due to water vapour feedback⁹: the warmer atmosphere contains more water vapour which in turn traps more radiation. The fact that $dG_a/dT > 6 \text{ W m}^{-2} \text{ K}^{-1}$ has important climate implications.

The surface emission increases at a rate close to $dE/dT = 4\sigma T^3 \approx 6.1 \text{ W m}^{-2} \text{ K}^{-1}$ at $\sim T = 300$ K. But as $dG_a/dT \geq dE/dT$, the energy cannot escape to space but instead is trapped in the atmosphere. Indeed, the slightly larger slope of G_a suggests that the trapping increases faster than the surface emission. The main deduction is that the warmer tropical-ocean/atmosphere system has lost the fundamental negative feedback between temperature and infrared emission which expels excess heat by radiating to space. This tropical 'super greenhouse effect' is caused by a combination of several mutually reinforcing factors, including increases in total-column H_2O ^{9,10,16}, H_2O continuum absorption¹⁷ which scales quadratically with H_2O partial pressure; higher middle- and upper-troposphere H_2O concentrations¹⁹, and changes in the lapse rate^{18,19}. The system is potentially unstable unless another negative feedback exists to stabilize it,

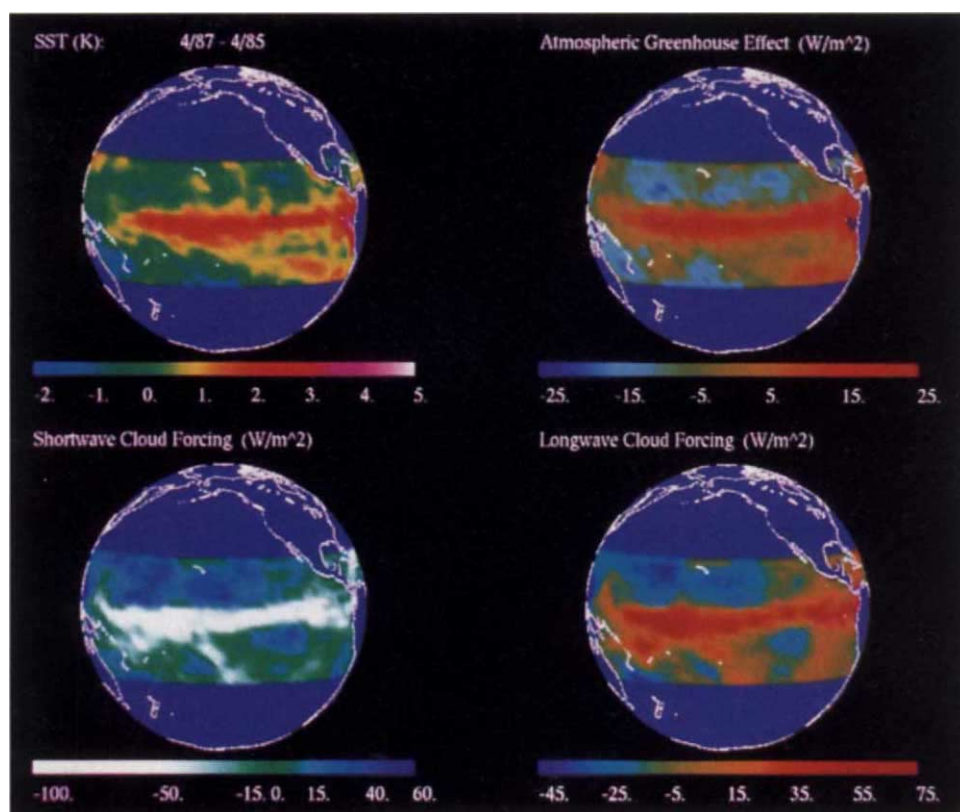


FIG. 1 Changes in the SST, the greenhouse effect and the cloud forcing accompanying the 1987 El Niño. Top left panel (a), April 1987 minus April 1985 monthly mean SSTs for 25° N to 25° S . Outside this region changes are much smaller. SST data are obtained from ref. 7. The equatorial warming began during late 1986 and was well established by April of 1987 (ref. 8). Top right panel (b), April 1987 minus April 1985 atmospheric greenhouse effect. Bottom left panel (c), April 1987 minus April 1985 cloud short-wave

forcing. The El Niño warming leads to highly reflective clouds which reduce (as $C_s < 0$) the solar energy absorbed by the oceans by as much as 100 W m^{-2} . Bottom right panel (d), April 1987 minus April 1985 cloud long-wave forcing, C_1 . The greenhouse effect of clouds increases significantly over the regions subject to the surface warming. Increased cirrus cloudiness in the upper troposphere is needed to account for changes in C_1 larger than 30 W m^{-2} .

and this brings us to the changes in the cloud forcing. The cloud-forcing terms C_1 (Fig. 1d) and C_s (Fig. 1c) also increase in magnitude for warmer oceans (Fig. 1a). A closer examination of the relationship between C_1 and C_s (Fig. 2c), reveals the strong degree of anti-correlation between the two, which can also be deduced from results published earlier²⁰⁻²². The generality of this finding is shown in Table 1. This remarkable correlation and the fact that large negative values of C_s are correlated with warmer waters (Fig. 1d) lead to the following conjecture: as the tropical oceans warm, the rapid rise of G_a with SST leads to an unstable feedback. The warming continues until the clouds become thick enough to shield the ocean from the solar radiation and arrest further warming. This thermostat effect of the cirrus anvils controls the maximum value of SSTs. The validity of this inference is examined next.

Experiment 2

In this experiment, we examined the flux changes that occurred during the 1987 El Niño. The observed changes are very similar to those inferred above. First, G_a increases with SST (Fig. 3a) at a rate of $6-9 \text{ W m}^{-2} \text{ K}^{-1}$. The plot of dC_s against dC_1 (Fig. 3b) confirms that clouds with large C_1 are also very reflective. As $dC_s/dG \geq -1$, the decrease in the absorbed solar energy nearly balances the increase in the total greenhouse effect (Fig. 3c). These results support the inference from experiment 1 that G increases significantly with a warming of the ocean, and this unstable increase is effectively offset by the brighter clouds resulting from the warming. Similar results are obtained for the months of February, May, July and October (Table 1), but results for February and July should be viewed with caution as we have not accounted for seasonal changes in the incoming solar radiation and the solar zenith angle, both of which have complex and largely unknown effects on the cloud-forcing terms.

To probe further into the link between the warming and the absorbed solar energy, we plot the product of dT and dC_s

against $(dT)^2$ for each $2.5^\circ \times 2.5^\circ$ region (Fig. 4). If the relationship is linear, we can obtain the rate of decrease of C_s with increase in SST from the slope, a_s . A statistically significant relationship does exist for April (Fig. 4 and the column under a_s in Table 1), although it is weaker for the other months; for example, for February 1987 minus February 1985 and for May 1987 minus May 1985, a_s is $\sim -15 \text{ W m}^{-2} \text{ K}^{-1}$. The correlation coefficient $|r|$ is only ~ 0.6 and is almost zero for July and October. But we may obtain a better picture by examining the relationship on a larger scale.

We focus on the region of maximum warming, the central and eastern Pacific from 180° (the date line) to 90° W and 5° N to 5° S . During April, for example, the average SST for this region increased from 300.1 in 1985 to 301.9 K in 1987 and the average C_s changed from -29 in 1985 to -63 W m^{-2} in 1987 ($dC_s = -34 \text{ W m}^{-2}$); for February, SST increases from 299.2 to 301 K and $dC_s = -32 \text{ W m}^{-2}$; for May, SST increases from 300 to 301.7 K and $dC_s = -34 \text{ W m}^{-2}$; for July, the changes are from 299 to 301 K and $dC_s = -12 \text{ W m}^{-2}$; and finally for October, SST increases from 298.6 to 300.3 K and $dC_s = -6 \text{ W m}^{-2}$. Several inferences can be made. First, on the larger scale the warming is accompanied by a decrease in solar absorption without exception. This is reassuring as the processes that link SST with C_s involve cooperative interaction between deep convection and the large-scale circulation which transports moisture into the region of warming from the rest of the tropical Pacific (details given later). Second, if we exclude the colder eastern Pacific from 90° W to 120° W , which is mostly devoid of deep convection, dC_s for October decreases from -6 to -13 W m^{-2} with similar decreases (ranging from -5 to -9 W m^{-2}) during other months. This is another indication of the role of deep convection in triggering the negative feedback between SST and C_s . Finally, the marked decrease of $-dC_s$ towards October, which roughly marks the final phase of the El Niño, is a reminder of the transient nature of the event. As the El Niño warming is

FIG. 2 Monthly mean greenhouse effect and cloud-forcing terms for the Pacific extending from 140° E to 90° W . Each point in this and other figures represents a monthly averaged value for a 2.5° (latitude) $\times 2.5^\circ$ (longitude) region. *a*, Total greenhouse effect, G , plotted as a function of SST for April 1987 from 85° N to 85° S . Regions with $\text{SST} < 274 \text{ K}$ are excluded. *b*, Atmospheric greenhouse effect G_a against SST for the tropical Pacific from 25° N to 25° S for April 1987. The least-squares linear slope is $dG_a/dT = 6.5 \text{ W m}^{-2} \text{ K}^{-1}$ with a correlation coefficient of $r = 0.96$. *c*, Cloud short-wave forcing against cloud long-wave forcing for the region 10° N to 10° S for 4/85(+), 2/87(○), 2/88(×), and 4/87(·). The linear correlation is $dC_s/dC_1 = -0.951$ ($r = -0.9$).

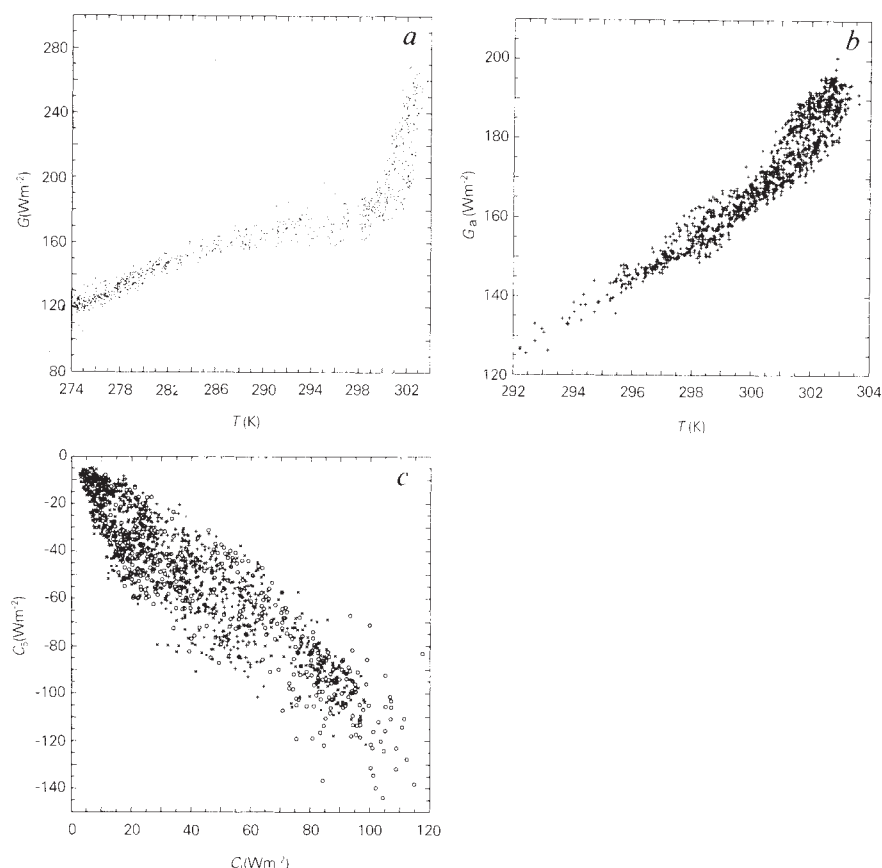
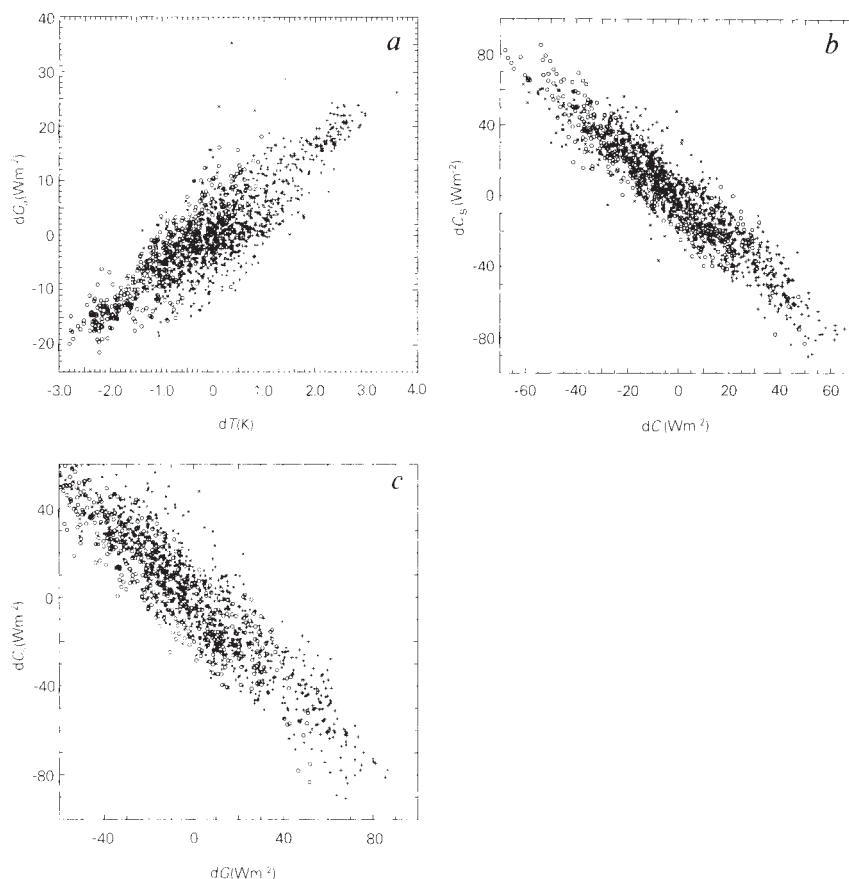


FIG. 3 Changes induced by 1987 El Niño in monthly mean greenhouse effect, cloud-forcing terms and SST for the Pacific ocean (140° E to 90° W) from 10° N to 10° S. The points represent the differences for 4/87 minus 4/85(+), 4/89 minus 4/87(○), and 2/88 minus 2/87(×). *a*, dG_a against dT where d denotes difference between two years. The linear fit is $dG_a/dT = 6.80 \text{ W m}^{-2} \text{ K}^{-1}$ ($r = 0.85$). *b*, Change in cloud short-wave forcing against change in cloud long-wave forcing. The linear fit is $dC_s/dC_l = -1.20$ ($r = -0.93$). As $dC_s/dC_l < -1$, clouds have a net cooling effect. *c*, Change in cloud short-wave forcing against change in total greenhouse effect $G = G_a + C_l$. The linear fit is $dC_s/dG = -0.92$ ($r = -0.92$), which indicates that the increase in the greenhouse effect is slightly larger than the decrease in solar absorption.



a transient phenomenon which started to revert back to 'normal' conditions around November 1987, considerable changes in the equatorial dynamics during the terminal phase of the El Niño¹³ can be expected to modify the radiative feedbacks.

To explore additional climate implications, we estimate the change of C_s with SST, although acknowledging that this is a hazardous task given the nonlinear nature of the coupling between SST, ocean-atmosphere dynamics and radiation physics. We can use two independent methods: (1) From Fig. 3*a*, $dG_a/dT \approx 6.8 \text{ W m}^{-2} \text{ K}^{-1}$. As $dC_s = -0.92 (dC_l + dG_a)$ (Fig. 3*c*) with the further constraint that $dC_s = -1.20 dC_l$ (Fig. 3*b*), we obtain $dC_s/dT \approx -27 \text{ W m}^{-2} \text{ K}^{-1}$ and $dC_l/dT \approx 23 \text{ W m}^{-2} \text{ K}^{-1}$. (2) The covariance plot (Fig. 4*a*) of dC_s and dT with $(dT)^2$ and the corresponding plot for dC_l (Fig. 4*b*) yield $dC_s/dT \approx -22$ and $dC_l/dT \approx 18$, where because of uncertainties in the SST data, regions with $dT < 1 \text{ K}$ are excluded.

Thermodynamics and dynamics of the feedbacks

In what follows, we describe the fundamental principles that govern the observed correlations described above. In the warm climate regimes of interest to this study, the thermodynamics of H_2O have a dominant influence on the radiative feedbacks. The saturation humidity, q_s , depends exponentially on temperature as $\{e^{(-5,400/T)}\}/T$ (T in kelvin). As a result, a 1% increase in T from $T = 300 \text{ K}$, for example, increases q_s by 17%. Indeed, satellite observations^{9,10} reveal that total atmospheric water vapour increases by $\sim 17\%$ per 1% increase in SST. Hence G_a increases with SST.

The latent energy of a parcel of air also grows by 17% per 1% increment in T . When $T > 300 \text{ K}$, a parcel of moist air near the surface has sufficient latent energy that if it is forcibly lifted until it reaches saturation, it can overcome the gravitational potential energy and rise to the upper troposphere as a cumulonimbus cloud^{11,12}. The ice crystals detrained by these deep clouds spread thousands of kilometres downwind in the

form of cirrus anvils^{22,23} and also moisten the upper troposphere¹⁸. The maxima in the cloud-forcing fields (Fig. 1) are caused by the cirrus anvils¹² and not by the cumulonimbus clouds, which are two orders of magnitude smaller²³. When SST drops below 300 K, the latent energy in the surface layer is generally not sufficient to form deep convection²⁴. Hence G increases rapidly when $T > 300 \text{ K}$.

Consider next the total static energy of the atmosphere, $h = C_p T + gZ + Lq$, where the first term is the thermal energy, the second is the gravitational potential energy and the third is the latent energy, with Z the altitude and q the H_2O mass mixing ratio. We rewrite it in the form $h/C_p = T + (g/C_p)Z + (L/C_p)q$ where $g/C_p = 9.8 \text{ K km}^{-1}$ and $L/C_p = 2,500 \text{ K}$. At $T = 303 \text{ K}$, the saturation q at the surface is 0.026, and assuming a reasonable value of 75% for the relative humidity, the latent energy is $\sim 50 \text{ K}$ and $h/C_p = 353 \text{ K}$. If we adopt the observed equatorial temperature and humidity profiles²⁵, h/C_p decreases from 353 K at $Z = 0 \text{ km}$ to 330 K at $Z = 6 \text{ km}$. Above 6 km it increases with Z to $\sim 350 \text{ K}$ at 16 km because of the dominance of the potential energy term. As a result, a surface parcel with SST = 303 K can penetrate to $\sim 16 \text{ km}$. When the SST drops to 297 K, the surface parcel has a static energy of only 330 K and can barely reach the middle troposphere.

The warmer ocean thus produces clouds at higher altitudes. As cloud-top temperature and emission of radiation to space decrease with altitude, these clouds trap more long-wave radiation and have a larger greenhouse effect²⁶. As a result C_l can increase with SST. Why does the solar reflection increase with SST? Deeper penetration by clouds is not sufficient to account for the enhanced solar reflection; in addition, the extent of cloudiness and the optical thickness of the clouds have to increase with SST. The required moisture for sustaining cirrus is not necessarily provided by local evaporation²⁷ but instead by large-scale transport within the lower troposphere into the region of convection^{27,28}. These large circulation systems are the

'Hadley' and 'Walker' circulations⁴. The sources of energy for these large-scale motions are the latent heat released by convection²⁹, the cirrus long-wave cloud forcing and the spatial gradients in SST³⁰. Therefore this convective large-scale system is self-sustaining. The large-scale convergence of moisture into the warm oceanic regions amplifies the warming through the enhanced atmospheric greenhouse effect, further driving the circulation. This continues until the cirrus clouds, which accumulate during this process, reflect enough sunlight to arrest further warming. Thus the anvils act like a thermostat.

Predictions of the thermostat hypothesis

The independent natural experiments support this hypothesis: the feedback between SST, convection and cloud reflection of solar radiation prevents a runaway greenhouse effect on the SST and acts as a thermostat to regulate the maximum ocean temperatures on the planet. The climate implications of this hypothesis are now developed in more detail.

Maximum ocean temperature. Let us consider an ocean region where dynamical transport of heat in the surface layers is negligible so that it can reach the maximum value imposed by the solar absorption. The equatorial western Pacific nearly satisfies this criterion because the net heat flux into the ocean is very small ($0\text{--}20\text{ W m}^{-2}$)^{31,32}. We can therefore apply equation (1) with a correction for atmospheric transport which is discussed next. An examination of the equatorial temperature gradients²⁵ reveals that the mid-to-upper troposphere has negligible east-west and north-south gradients in spite of the fact that the latent heating ($200\text{--}300\text{ W m}^{-2}$)^{4,31} and cloud greenhouse effect add as much as $300\text{--}400\text{ W m}^{-2}$ heat flux in the western portions of the Pacific. Seasonal and longitudinal average temperatures in the upper troposphere differ by less than 2 K between 25° N to 25° S . Likewise, the upper troposphere in the equatorial western Pacific is less than 2 K warmer than that of the eastern Pacific 10^4 km away. This homogeneity follows from the balancing of latent and radiative heating in the tropical atmosphere by adiabatic cooling resulting from convective and large-scale overturning³³. In particular, adding energy to the upper troposphere is very effective in driving the Walker and the Hadley cell circulation²⁹. Recent general circulation model studies^{34,35} also suggest that the radiative energy converging in the anvils does not cause localized warming over the anvil regions. Instead, the warming is uniformly distributed over the entire tropical region.

We therefore assume that the C_1 term, which is largely heat trapped in the troposphere, is transported out of the western Pacific, and only a small fraction contributes to warming the ocean locally. This fraction, f , ranges between 0 and 0.2. The upper bound is related to the fraction of the long-wave emission from the base of cirrus clouds that can reach the ocean surface directly. Incidentally, C_1 ($\sim 60\text{--}80\text{ W m}^{-2}$) is nearly equal to the net annual mean radiative heating of 80 W m^{-2} (ref. 36) at the

top of the atmosphere minus the oceanic heat transport³² of $\sim 0\text{--}20\text{ W m}^{-2}$ in the western Pacific region. This implies that 60 W m^{-2} is being transported out of the region by advection within the atmosphere, which is consistent with the assumed range for f .

On the other hand, the C_s term represents a direct loss of solar energy to the ocean surface³⁴. With the above inferences, equation (1) can be rephrased as follows:

$$\text{SST}_{\max} = T_0 + \{(S_c + G_0 - \sigma T_0^4)/\beta\} \quad (2)$$

$$\text{where } \beta = 4\sigma T_0^3 - dG_a/dT - f dC_1/dT - dC_s/dT$$

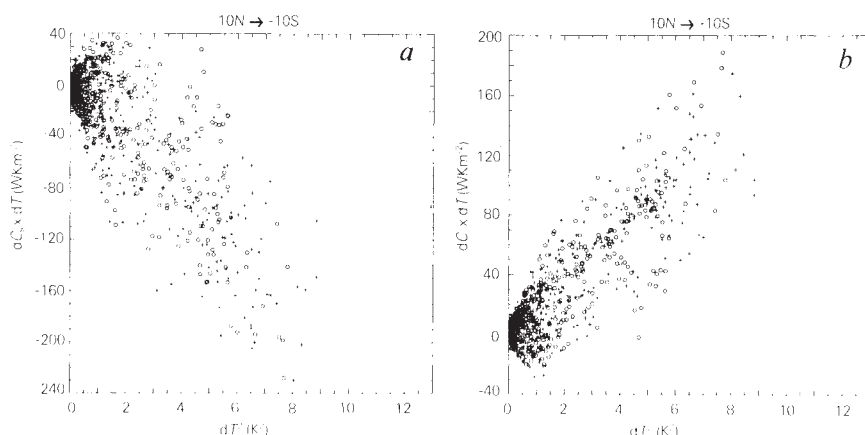
The net radiative feedback is denoted by β . The variables have been expanded about a reference temperature T_0 ; for example, $G_a = G_0 + dG_a/dT$, where $G_0 = 165\text{ W m}^{-2}$ is the value of G_a at T_0 , assumed to be 300 K. It is safe to assume an uncertainty of $\sim \pm 15\text{ W m}^{-2}$ in the radiative flux terms. The clear-sky solar absorption is $S_c = S(1 - A_c) = 370\text{ W m}^{-2}$, where $S = 424\text{ W m}^{-2}$ is the annual mean equatorial solar radiation and $A_c = 0.1$ is the clear-sky albedo, and we subtract 12 W m^{-2} for stratospheric solar absorption³⁷. Similarly, the stratospheric contribution of 5 W m^{-2} to G_a has been subtracted from G_0 . The values derived from Figs 3 and 4 are adopted for the various feedback terms in β . Changes in the equatorial atmospheric temperature and humidity profiles in response to SST changes are implicitly included in the long-wave feedback terms, as dG_a and dC_1 are obtained directly from observed changes. The cloud feedback terms arising from convection will be zero when $T < T_0$, and the black-body term is $4\sigma T_0^3 = 6.1\text{ W m}^{-2}\text{ K}^{-1}$. Note first that in the absence of the cloud feedback terms, the black-body negative feedback (the first term in the denominator) is dominated by the positive feedback effect of dG_a/dT ($6.8\text{ W m}^{-2}\text{ K}^{-1}$), creating a potentially unstable state. Substitution of $-27 < dC_s/dT < -22$ and $18 < dC_1/dT < 23$ yields:

$$303 < \text{SST}_{\max} < 305\text{ K}$$

The lower limit is for $f = 0.0$ and $dC_s/dT = -27$, and the upper limit of 305 K is for $f = 0.2$ and $dC_s/dT = 22\text{ W m}^{-2}\text{ K}^{-1}$ (from Fig. 3). If we allow 20 W m^{-2} for oceanic heat transport, the 305 K value drops to $\sim 304\text{ K}$ and the 303 K value drops to 302 K in agreement with observed western Pacific SSTs (Fig. 2a). For the above example, the feedback value ranges between $15 < \beta < 26\text{ W m}^{-2}\text{ K}^{-1}$. This large damping should tend to minimize the seasonal variations of the SST in the western Pacific.

Evaporative cooling of the warm oceans is another possible mechanism^{15,25} for limiting SSTs, but this cannot be a limiting factor for several reasons: (1) the added moisture to the boundary layer will enhance G_a ; and (2) observed precipitation exceeds evaporation by nearly a factor of 2 in the western Pacific^{4,25,31} (the western Pacific ocean imports rather than

FIG. 4 a, Change induced by 1987 El Niño in solar absorption by the ocean-atmosphere system. Product of dC_s and dT against $(dT)^2$ for the Pacific, 10° N to 10° S . The points represent the differences between 4/87 and 4/85(+) and 4/89 and 4/87(○). A negative value for the product indicates a decrease (increase) in absorbed solar energy with a warming (cooling) of the ocean. The linear fit is $dC_s/dT = -19\text{ W m}^{-2}\text{ K}^{-1}$ ($r = -0.80$). The error in SST data is 1 K, and when regions with $dT < 1\text{ K}$ are excluded the fit is $dC_s/dT = -22\text{ W m}^{-2}\text{ K}^{-1}$ ($r = -0.73$). b, Same as (a) but for cloud long-wave forcing. The linear fit is $dC_1/dT \approx 16$ and increases to $17.5\text{ W m}^{-2}\text{ K}^{-1}$ when points with $dT < 1\text{ K}$ are excluded.



exports water vapour). This inference, albeit based on imprecise observations, is also substantiated by the low salinity of the surface water³¹. What is needed is a permanent loss of energy, and reflection of sunlight back to space is the most efficient mechanism to accomplish this task.

El Niño warming. The basic hypothesis should apply during an El Niño as well. Indeed, during both the 1987 El Niño and the more severe 1983 El Niño⁴, the maximum SST did not exceed 303 K. In the central and eastern Pacific, upwelling of colder water from below and advection of cold water from the eastern coast causes SSTs to drop below 303 K (refs 4, 31). When the dynamical cooling effects abate during periods such as an El Niño year, the SST approaches the maximum value permitted by the anvil thermostat. Thus the El Niño year constitutes the equilibrium state of the Pacific, whereas the non El Niño years are the periods when the central and east Pacific is pulled away from equilibrium by the ocean dynamics.

Greenhouse and solar-dominated climate regimes

For the oceans⁹, the globally averaged greenhouse effect (G) is 175 W m^{-2} and the absorbed solar radiation is 240 W m^{-2} . For the warmest equatorial oceans, the maximum values are $G \approx 280$ (Fig. 2a) and solar absorption $\sim 300 \text{ W m}^{-2}$ (ref. 36). In the absence of clouds, S can be as large as 380 W m^{-2} (ref. 6). Thus as the planet gets warmer, it shifts from a solar-energy-dominated regime to a regime in which the greenhouse effect becomes comparable to the solar absorption. The extreme example of this convective response to warming is Venus, with a convectively driven deep troposphere of $\sim 60 \text{ km}$, overcast conditions

and a high planetary albedo of 80% (compared with 30% for Earth). As a result, although Venus receives twice as much solar radiation as Earth at the top of the atmosphere, the absorbed solar radiation is smaller ($\sim 150 \text{ W m}^{-2}$). But in spite of this brightening effect, the planet is hotter (750 K) because its atmospheric greenhouse effect exceeds $17,000 \text{ W m}^{-2}$.

The feedback between convection, greenhouse effect and planetary brightness limits the temperatures in the warmest regions of the Earth to less than 305 K. What is the implication of this negative feedback if its validity extends to a perturbed atmosphere? It would take more than an order-of-magnitude increase in atmospheric CO_2 to increase the maximum SST by a few degrees, in spite of a significant warming outside the equatorial regions. In this regard, the present hypothesis departs considerably from modern general circulation models. It is encouraging, however, that some model studies^{1,16} are finding a marked feedback in the thermodynamic link between SST and cirrus properties. More significantly, a recent coupled ocean-atmosphere/general circulation model study²⁸ has suggested that the interaction between SST and solar radiation at the ocean surface is an important negative feedback in the El Niño dynamics. This result was also indicated qualitatively by earlier studies^{15,38}. For the hypothesis to become a validated theory, we need to establish the link between convection dynamics, large-scale moisture convergence into the warm regions and the brightness of those regions. Three-dimensional climate models in conjunction with detailed field and satellite observations can provide that link, but first the models must pass the test posed by the El Niño observations. □

Received 26 February; accepted 5 April 1991.

- Mitchell, J. F. B., Senior, C. A. & Ingram, W. J. *Nature* **341**, 132–134 (1989).
- IPCC *Climate Change* (Cambridge University Press, 1990).
- Cess, R. D. *et al. Science* **245**, 513–516 (1989).
- Philander, S. G. *El Niño, La Niña & the Southern Oscillation* (Academic, London, 1990).
- Arduin, P. E. & Kyle, H. L. *Mon. Weath. Rev.* **114**, 415–433 (1986).
- Harrison, E. F. *et al. J. geophys. Res.* **95**, 18687–18703 (1990).
- Reynolds, R. W. *J. Clim.* **1**, 75–86 (1988).
- WMO Report: *The Global Climate System 1986–1988 3rd Biennial Rev.* (CSMR 89/86 Geneva).
- Raval, A. & Ramanathan, V. *Nature* **342**, 758–761 (1989).
- Stephens, G. L. *J. Clim.* **3**, 634–645 (1990).
- Betts, A. K. & Ridgway, W. J. *atmos. Sci.* **46**, 2621–2641 (1989).
- Houze, R. A. & Betts, A. K. *Rev. Geophys. Space Phys.* **19**, 541–576 (1981).
- Cane, M. A. & Zebiak, S. E. *Science* **228**, 1085–1087 (1985).
- Gadgil, S., Joseph, P. V. & Joshi, N. V. *Nature* **312**, 141–143 (1984).
- Graham, N. E. & Barnett, T. P. *Science* **238**, 657–659 (1987).
- Somerville, R. C. & Remer, L. J. *geophys. Res.* **89**, 9668–9672 (1984).
- Hallberg, R. thesis, Univ. of Chicago (1990).
- Rind, D. *et al. Nature* **349**, 500–503 (1991).
- Hallberg, R. Eureka Conf. Abstr. ENV1.1K CALTECH (Calif. Inst. Technol., Pasadena, California, 1991).
- Cess, R. D., Briegleb, B. P. & Lian, M. S. *J. Atmos. Sci.* **39**, 53–59 (1982).
- Randel, D. L. & von der Haar, T. H. *J. Clim.* **3**, 1168–1173 (1990).

- Liou, K. *Mon. weath. Rev.* **114**, 1167–1199 (1986).
- Houze, R. A. & Hobbs, P. V. *Adv. Geophys.* **24**, 225–314 (1982).
- Fu, R., Del Genio, A. D. & Rossow, W. J. *Clim.* **3**, 1129–1152 (1990).
- Newell, R. E. *et al. The General Circulation of the Tropical Atmosphere* Vol. 1 (MIT Press, Massachusetts, 1974).
- Ramanathan, V. *J. Atmos. Sci.* **34**, 1886–1897 (1977).
- Cornejo-Garrido, A. G. & Stone, P. H. *J. Atmos. Sci.* **34**, 1155–1162 (1977).
- Barnett, T. P. *et al. J. Clim.* (in the press).
- Hartmann, D. L. *et al. J. Atmos. Sci.* **41**, 113–121 (1984).
- Lindzen, R. S. & Nigam, S. *J. Atmos. Sci.* **44**, 2418–2428 (1987).
- Pickard, G. L. & Emery, W. J. *Descriptive Physical Oceanography* (Pergamon, Oxford, 1982).
- Gent, P. R. *J. geophys. Res.* **96**, 3323–3330 (1991).
- Holton, J. *Introduction to Dynamic Meteorology* (Academic, 1972).
- Slingo, A. & Slingo, J. M. Q. *J. R. Met. Soc.* **114**, 1027–1062 (1988).
- Ramaswamy, V. & Ramanathan, V. *J. Atmos. Sci.* **46**, 2293–2310 (1989).
- Stephens, G. *et al. J. geophys. Res.* **86**, 9739–9760 (1981).
- Ramanathan, V. & Dickinson, R. E. *J. Atmos. Sci.* **36**, 1084–1104 (1979).
- Ramanathan, V. *J. Met. Soc. Japan spec. vol. Short- and Medium-Range Numerical Weather Prediction*, 151–176 (1987).

ACKNOWLEDGEMENTS. We thank the NASA Langley Research Center for processing the calibrated ERBE data, and A. Heymsfield and S. Sherwood for their comments on the paper. This work is supported by NASA, NSF and the Vetlesen fund.

Pathway of phosphatidylinositol(3,4,5)-trisphosphate synthesis in activated neutrophils

L. R. Stephens, K. T. Hughes* & R. F. Irvine

Biochemistry Department, AFRC Institute of Animal Physiology & Genetics Research, Cambridge Research Station, Babraham Hall, Cambridge CB2 4AT, UK

* Amersham International, Whitechurch, Cardiff CF4 7YT, UK

Neutrophils activated by the formyl peptide f-Met-LeuPhe transiently accumulate a small subset of highly polar inositol lipids. A similar family of lipids also appear in many other cells in response to a range of growth factors and activated oncogenes, and are presumed to be the direct or indirect products of 3-phosphatidylinositol kinase. The structures of these lipids are shown to be phosphatidylinositol 3-phosphate, phosphatidylinositol-(3,4)bisphosphate and phosphatidylinositol-(3,4,5)trisphosphate, and we present evidence that in intact neutrophils a phosphatidylinositol-(4,5)bisphosphate-3-kinase seems to be the focal point through which agonists stimulate the formation of 3-phosphorylated inositol lipids.

SEVERAL extracellular agonists and growth factors stimulate the rapid transient formation of at least two highly polar inositol phospholipids¹⁻⁶, neither of which seem to possess the conventional head-group structure of phosphatidylinositol(4,5)-bisphosphate, the ubiquitous precursor of the second messenger $\text{Ins}(1,4,5)\text{P}_3$ (ref. 7). The molecules are of interest not only because they represent a new class of compounds, rapidly generated on cell activation, but also because their synthesis correlates, possibly causally, with the agonist-driven entry of some cells into the mitotic cycle^{8,9}. Moreover, an enzyme activity that can synthesize chromatographically similar compounds *in vitro*³ is translocated rapidly in the cell when activated by many growth factor receptors^{3,4,10} and on transformation of cells by oncogenes¹¹⁻¹³. Although the structures of these new lipids are likely to be the 3-phosphorylated analogues of $\text{PtdIns}(4)\text{P}$ and $\text{PtdIns}(4,5)\text{P}_2$ ¹³⁻¹⁵, this has not been directly established, and moreover recent reviews^{14,15} have concluded that very different pathways are responsible for their appearance in agonist-stimulated cells (Fig. 5).

Polyphosphoinositides in human neutrophils

Traynor-Kaplan *et al.*^{1,2} showed that in human neutrophils *N*-formylmethionylleucylphenylalanine (FMLP) causes the rapid accumulation of two novel inositol phospholipids which are likely to be a PtdInsP_2 (although not $\text{PtdIns}(4,5)\text{P}_2$) and a PtdInsP_3 . We have similarly labelled intact human neutrophils with [³H]Ins and/or ³²Pi *in vitro*, and then analysed their head-group structures and metabolism. Two PtdInsPs , two PtdInsP_2 s and a PtdInsP_3 were characterized. The structures of three polyphosphoinositol lipids, phosphatidylinositol 4-phosphate ($\text{PtdIns}(4)\text{P}$), phosphatidylinositol 3-phosphate ($\text{PtdIns}(3)\text{P}$) and phosphatidylinositol(4,5)-bisphosphate ($\text{PtdIns}(4,5)\text{P}_2$), from bovine brain¹⁶ and cultured astrocytoma cells¹⁷, have been established, and are confirmed here for human neutrophils (data not shown). The two remaining lipids (which on deacylation

yielded peaks 1 and 3 in Fig. 1) had the structures of D-phosphatidylinositol(3,4)-bisphosphate ($\text{PtdIns}(3,4)\text{P}_2$) and D-phosphatidylinositol(3,4,5)trisphosphate ($\text{PtdIns}(3,4,5)\text{P}_3$). These structures were deduced by enzymatic and chemical dissection and a series of high-resolution chromatographic techniques to resolve the products in the presence of relevant internal standards (Figs 1 and 2).

Synthesis of polyphosphoinositides in neutrophils

The structures suggest, but do not define, the route(s) by which the phospholipids are constructed in the cell (for example $\text{PtdIns}(3,4,5)\text{P}_3$ could be derived by the phosphorylation of either $\text{PtdIns}(3,4)\text{P}_2$ or $\text{PtdIns}(4,5)\text{P}_2$). Clues may be gained about these processes by studying cell extracts capable of catalysing the interconversion of the various metabolites. But such results must be interpreted with caution, because in the intact cell a huge variety of compartmentation factors or 'environmental' conditions not mimicked *in vitro* may alter the distribution, pattern or even direction of flows of intermediates through the pathways.

Using intact human neutrophils, we have analysed the relative rates at which ³²Pi is incorporated into each of the phosphate groups of the polyphosphoinositides *in vitro*. ³²Pi will be incorporated most rapidly into a polyphosphoinositide if it is the last phosphate moiety to be added and progressively less rapidly earlier in the pathway (assuming each of the phosphate moieties comes from the same precursor, for example the γ -phosphate of ATP). After prolonged periods of incubation, when the system approaches equilibrium, these phosphate-specific differences in labelling vanish (for a more detailed explanation of this strategy, see refs 18, 19).

We labelled the cells with ³²Pi for 18 min and quenched them after 8 seconds in the presence of FMLP. This allowed the label to enter the precursors but was sufficiently brief that the specific radioactivities of the γ -phosphate of ATP and the phosphate groups of the polyphosphoinositides, and also the concentration of $\text{PtdIns}(3,4,5)\text{P}_3$, were still increasing rapidly (and moreover that the 5-phosphate of $\text{PtdIns}(4,5)\text{P}_2$ was significantly more radioactive than the 4-phosphate; L.R.S., unpublished data, and see Table 1 and Fig. 3). The distribution of ³²P within the various polyphosphoinositide species was then determined as described in the legend to Table 1.

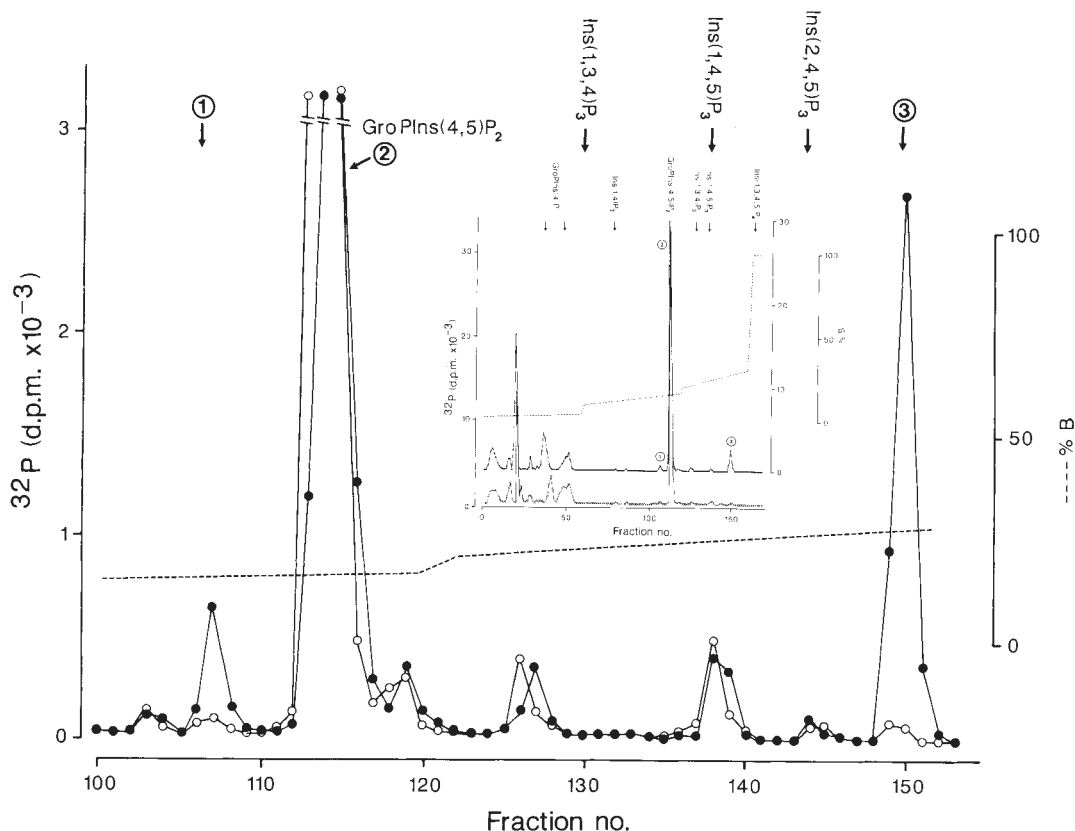
Of the total ³²P in $\text{PtdIns}(4,5)\text{P}_2$, 0.2, 43 and 57% were located in the 1- (phosphodiester), 4- and 5- (phosphomonoester) phosphates, respectively. Similarly, essentially all of the total ³²P in $\text{PtdIns}4\text{P}$ was in the 4-phosphate. These results confirm the metabolic map defined by Brockerhoff and Ballou¹⁸ ($\text{PtdIns} \rightleftharpoons \text{PtdIns}(4)\text{P} \rightleftharpoons \text{PtdIns}(4,5)\text{P}_2$) and represent the only direct evidence that in an intact hormone-sensitive cell, $\text{PtdIns}(4,5)\text{P}_2$ is synthesized using $\text{PtdIns}(4)\text{P}$.

A similar analysis of $\text{PtdIns}(3,4,5)\text{P}_3$ (Table 1) revealed that the 3-phosphate contained the largest share of ³²P, suggesting that this lipid had been formed by the phosphorylation of $\text{PtdIns}(4,5)\text{P}_2$. The distribution of ³²P among the 1-, 4- and 5-phosphates of $\text{PtdIns}(4,5)\text{P}_2$ and $\text{PtdIns}(3,4,5)\text{P}_3$ was consistent with the synthesis of $\text{PtdIns}(3,4,5)\text{P}_3$ beginning with

FIG. 1 Anion-exchange HPLC separation of the deacylation products derived from lipid extracts of either control (○) or FMLP-stimulated (●) ^{32}P -labelled (8 s) human neutrophils. Inset figure shows the entire elution profile. The elution times of internal ^{32}P -labelled standards are shown. The traces are typical of over 60 separations. Internal GroPIns(4)P standards usually eluted as two poorly defined peaks (see below and the legend to Fig. 3).

METHODS. A partispher 5-SAX column was eluted at 1 ml min^{-1} with a gradient of $2.5\text{ M NaH}_2\text{PO}_4$ (pH 3.8 with NaOH (%B)); fractions were collected every 0.5 min. Human neutrophils were labelled *in vitro* with either (1) ^3H -Ins or (2) ^{32}P , as follows. (1) Three separate 5-ml incubations, each containing 5×10^7 cells per ml (see legend to Fig. 3) and 1 mCi ml^{-1} ^3H -Ins (Amersham; 18 Ci mmol^{-1}) in RPMI medium (without Ins; Gibco) supplemented with

1% (v/v) penicillin and streptomycin (Flow Labs), 10% (v/v) 190 mM HEPES (pH 7.2, 37°C with NaOH), 1% (v/v) 200 mM EGTA (pH 7.2, 37°C with NaOH) and 0.5% (w/v) BSA. The incubations were shaken in an orbital mixer at 200 r.p.m. and 37°C for 5.5 h, then stimulated with FMLP ($1\text{ }\mu\text{M}$ final concentration) for 15 s before quenching the cells and extracting their phospholipids as described in the legend to Fig. 3b. To one of the three incubations, $0.1\text{ }\mu\text{Ci}$ of ^{32}P -PtdIns(4,5) P_2 , prepared as described³¹, was added. (2) A 1.3-ml incubation containing 5×10^7 cells per ml and 20 mCi ml^{-1} ^{32}P (PBS-43, Amersham) as described in the legend to Fig. 3, except that the cells were labelled for only 18 min before being stimulated with FMLP ($1\text{ }\mu\text{M}$) for 8 s with the original ^{32}P , still present in the incubation. The cells were quenched and the phospholipids extracted as described above. The lipid extracts were deacylated and resolved by anion-exchange HPLC as described above and in Fig. 3. From the labelling scheme (1) three ^3H -labelled peaks were collected (their ^{32}P -labelled counterparts are illustrated) which eluted as follows. 1, A peak just before the internal ^{32}P -GroPIns(4,5) P_2 (derived from the ^{32}P -PtdIns(4,5) P_2 spiked into one of the three parallel extracts) which was agonist-sensitive and contained no ^{32}P . 2, The major ^3H -Ins-labelled peak that co-migrated with the internal ^{32}P -GroPIns(4,5) P_2 generated by deacylation of the added ^{32}P -PtdIns(4,5) P_2 . 3, A peak that ran at the time expected for the agonist-sensitive 'GroPIns P_3 ' studied (and previously described^{1,2}), which contained no ^{32}P from the internal ^{32}P -PtdIns(4,5) P_2 and co-migrated with the water-soluble ^{32}P -labelled deacylation product of the ^{32}P -lipid formed in assays containing PtdIns(4,5) P_2 , γ - ^{32}P -ATP and proteins immunoprecipitated with an anti-phosphotyrosine monoclonal antibody (ICN) from PDGF-stimulated Swiss 3T3 cells³² (data not shown). These three peaks were isolated, pooled (from all three samples in the cases of peaks 1 and 3 and only from those two ^3H -labelled extracts which were not spiked with ^{32}P -PtdIns(4,5) P_2 in the case of 3) and desalted¹⁸. The total ^3H d.p.m. recovered in compounds 1, 2 and 3 were 97,700 d.p.m., 2,148,250 d.p.m. and 210,000 d.p.m., respectively. From the labelling scheme (2), three ^{32}P -labelled peaks showing the same properties as the three ^3H -Ins-labelled peaks defined above (and which contained 26,487, 773,934 and 47,558 d.p.m. of ^{32}P , respectively) were similarly isolated from the ^{32}P -labelled extract. Aliquots (50–70%) from each of the three ^3H -Ins-labelled



peaks were mixed with all of these corresponding ^{32}P -labelled peaks (except in the case of ^{32}P -GroPIns(4,5) P_2 , of which only 20% was used) then re-chromatographed on an anion-exchange HPLC column as described. In each case the ^{32}P and ^3H co-eluted as single peaks at the same time as they had originally, and these were collected and desalted before further analysis (Table 1). Aliquots of the residual ^3H -labelled compounds were analysed as follows to determine their structure. Peaks 1, 2 and 3 all gave ^3H -GroPIns when dephosphorylated with alkaline phosphatase¹⁷. Peak 1 was deglycerated¹⁷, the product co-chromatographed with internal ^{32}P -Ins(1,3,4) P_3 but not with internal ^{32}P -Ins(1,2,4) P_3 (Fig. 2d) and on periodate oxidation, reduction and dephosphorylation¹⁶ (PRD) yielded ^3H -L-altritol (refs 33, 34). This compound was therefore Ins(1,3,4) P_3 and as its 1-phosphate and the diester phosphate of peak 1 are identical (they both contained $<0.5\%$ of the total ^{32}P in their respective molecules, see Table 1 and ref. 17), peak 1 must be GroPIns(3,4) P_2 . Deglyceration of peak 2 yielded a compound which co-chromatographed with internal ^{32}P -Ins(1,4,5) P_3 but not internal ^{32}P -Ins(1,4,6) P_3 (similar to Fig. 2c) and upon PRD yielded ^3H -D-iditol. This compound is therefore Ins(1,4,5) P_3 and by the same logic as above (Table 1) its 1-phosphate and the diester phosphate of 2 are equivalent, so peak 2 is GroPIns(4,5) P_2 . Peak 3 was deglycerated to an Ins P_4 whose chromatographic properties are illustrated in Fig. 2a, b, and upon PRD gave ^3H -inositol (unless pretreated with 1 M HCl at 80°C for 8 min, in which case it gave ^3H -glucitol and ^3H -inositol). The Ins P_4 was also dephosphorylated with human erythrocyte plasma membranes (5-ml assays containing 4 mg ml^{-1} erythrocyte protein, 12.5 mM HEPES, 1.0 mM EGTA, 1 mg ml^{-1} BSA, pH 7.0, 37°C and 500 d.p.m. of ^{32}P -Ins(1,3,4,5) P_4) in the presence of either 10 mM MgCl_2 for 30 min or of 5 mM EDTA for 60 min. Under both sets of conditions the ^3H -Ins P_4 was metabolized at an identical rate to the internal ^{32}P -Ins(1,3,4,5) P_4 . The former conditions gave an ^3H -Ins P_3 which was identified as ^3H -Ins(1,3,4) P_3 by the same criteria as above (similar to Fig. 2d) and the latter, ^3H -Ins(1,4,5) P_3 , again identified by the same criteria as above (Fig. 2c). As with peaks 1 and 2 above, the data in Table 1 establish that the 1-phosphate of the Ins(1,3,4,5) P_4 derived from peak 3 was the only phosphate moiety that contained an equivalent amount of ^{32}P to the phosphodiester phosphate of 3 and hence peak 3 is GroPIns(3,4,5) P_3 .

TABLE 1 ^{32}P distribution in polyphosphoinositides extracted from FMLP-stimulated neutrophils

Phospholipid	Phosphodiester phosphate	Identity of individual phosphates (% of total ^{32}P)				100% values total d.p.m.)	
		1	3	4	5	^3H	^{32}P
PtdIns(3)P	0.0	0.2	100.7			5,508	1,906
PtdIns(4)P	0.2	0.4		99.6		95,020	50,125
PtdIns(3,4)P ₂	0.3	0.5 ± 0.67 (3)	62.0 ± 0.3 (2)	37.8 ± 1.2 (2)		54,429	22,463
PtdIns(4,5)P ₂	0.0	0.2		43.2	56.6	1,565,731	88,135
PtdIns(3,4,5)P ₃	0.2	0.04 ± 0.8 (4)	44.8 ± 0.7 (5)	24.9 ± 0.8 (4)	30.4 ± 0.6 (2)	110,991	38,858

Human neutrophils were labelled with ^{32}P for 18 min, stimulated with FMLP for 8 s, quenched and a phospholipid extract deacylated as described in the legend to Fig. 1 (the glycerophosphoinositol phosphates GroPIns(3)P and GroPIns(4)P analysed in these experiments were obtained from samples prepared in parallel to those described in Fig. 1, except that they were purified by TLC before deacylation; see legend to Fig. 3). The intramolecular distribution of ^{32}P in each of the polyphosphoinositides was determined as described below and is shown as the percentage of the total ^{32}P in a particular lipid found in each of its constituent phosphate moieties. Where more than one independent estimate (by analysis of different degradation products, see below) of these values was obtained (number defined in brackets), then the figures are given as means ± the maximum range of the data used to define that mean. The total ^3H and ^{32}P d.p.m. which defined the 100% values for each of the lipids are indicated in the table. The overall strategy and liquid scintillation counting techniques used were as described in ref. 18. The starting compounds (dual-labelled with ^3H in the inositol moiety, and ^{32}P) were systematically dephosphorylated, the products at each stage were purified by HPLC, fully structurally characterized and their ^3H and ^{32}P content quantified. The ^3H recovered served as a measure of the yield and the ^{32}P content defined the total ^{32}P contained by the phosphates present in that particular metabolite. The ^{32}P content of individual phosphate moieties was determined as differences in the molar ^{32}P content of the various metabolites, or directly when an inositol monophosphate was considered. The ^{32}P content of the phosphodiester phosphate of each compound was defined by incubating their respective glycerophosphoinositol phosphates with alkaline phosphatase¹⁷. This showed that the phosphodiester phosphates in the glycerophosphoinositol phosphates and the 1-phosphates in the inositol phosphates derived from them were structurally equivalent (see Fig. 1). The remaining samples were converted to their corresponding inositol phosphates¹⁷ (this process had no effect on the $^3\text{H}/^{32}\text{P}$ content of any of the lipid derivatives being analysed) and these were then enzymatically dephosphorylated in various ways. Ins(1,3,4,5)P₄ samples were metabolized with human erythrocyte plasma membranes either in the presence of Mg^{2+} or EDTA (see legend to Fig. 1) to form Ins(1,3,4)P₃ and Ins(1,4,5)P₃, respectively. Ins(1,3,4)P₃ samples (either derived from GroPIns(3,4)P₂ or via Ins(1,3,4,5)P₄ from GroPIns(3,4,5)P₃) were metabolized with (1) Ins(1,4)P₂–Ins(1,3,4)P₃ 1-phosphate phosphatase (purified 800-fold from bovine brain supernatant⁴⁸ by ammonium sulphate fractionation and anion-exchange chromatography on a Mono Q FLPC column^{49,50}) to form P_i and Ins(3,4)P₂, or (2) a bovine brain supernatant (0.55 mg ml⁻¹ final protein in the presence of 5 mM EDTA, 25 mM HEPES, 1 mM EGTA, pH 7.0, 37 °C) to form P_i, Ins(1)P and Ins(1,3)P₂. Ins(1,4,5)P₃ (either derived from GroPIns(4,5)P₂ or via Ins(1,3,4,5)P₄ from GroPIns(3,4,5)P₃) was metabolized with either (1) human erythrocyte ghosts in the presence of Mg^{2+} (see legend to Fig. 1) to form Ins(1,4)P₂, or (2) a bovine brain supernatant (2 mg ml⁻¹ final protein in the presence of 5 mM MgCl_2 , 5 mM Ins(2)P, 25 mM HEPES, 1 mM EGTA, pH 7.0, 37 °C) to form predominantly Ins(4)P (over 90% of the ^3H -labelled products was recovered in Ins(4)P). Ins(1,3)P₂ (derived from GroPIns(3)P) was converted to Ins(1)P and P_i as described¹⁷. Ins(1,4)P₂ (derived from GroPIns(4)P) was metabolized to Ins(4)P under the conditions defined for Ins(1,4,5)P₃ above. Ins(3,4)P₂ (derived from GroPIns(3,4)P₂ through Ins(1,3,4)P₃ or from GroPIns(3,4,5)P₃ through Ins(1,3,4,5)P₄ and Ins(1,3,4)P₃) was metabolized with bovine brain supernatant (under the conditions defined for the metabolism of Ins(1,3,4)P₃ above); the only significant products were Ins(3)P and P_i.

PtdIns and progressing through the intermediates PtdIns(4)P and PtdIns(4,5)P₂. The data are neither consistent with the synthesis of PtdIns(3,4,5)P₃ by the 5-phosphorylation of PtdIns(3,4)P₂ (Fig. 5c and below), nor with 4-phosphorylation of PtdIns(3,5)P₂ (a lipid recently found in some cells^{13,14}). The absence of any detectable PtdInsP₄ in unstimulated cells (or FMLP-stimulated neutrophils) and the rapid accumulation of PtdInsP₃ in response to FMLP (Fig. 3) make it unlikely that PtdIns(3,4,5)P₃ arises by a dephosphorylation.

Analysis of PtdIns(3,4)P₂ extracted from human neutrophils labelled with ^{32}P and stimulated with FMLP (Fig. 1 and Table 1) showed that its 3-phosphate had the highest specific radioactivity and its 1-phosphate the lowest. This implies that PtdIns(3,4)P₂ could be synthesized by the phosphorylation of PtdIns(4)P (but not by the phosphorylation of PtdIns(3)P). This intramolecular distribution of ^{32}P would, however, also arise if PtdIns(3,4)P₂ was formed by the removal of the 5-phosphate from PtdIns(3,4,5)P₃.

A similar situation emerges for the ^{32}P -labelling of PtdIns(4)P (see above) and PtdIns(3)P; both contained over 99% of their total ^{32}P in their monoester phosphate groups. This simple pattern is compatible with the *de novo* synthesis of each of these compounds from PtdIns. But it is possible (given the existence of two forms of PtdInsP₂ and also a PtdInsP₃ containing both 3 and 4 monoester phosphate moieties) that one (or both) of these lipids is formed by dephosphorylation of more highly phosphorylated lipids. We know that PtdIns(4)P can be formed by both mechanisms in reactions catalysed by PtdIns(4)-OH kinase and PtdIns(4,5)P₂ 5-phosphatase^{20–22}. These alternative explanations for the labelling patterns of PtdIns(3)P and

PtdIns(3,4)P₂ are not easily distinguished by experiments using intact cells, and to resolve the issue two strategies can be adopted: (1) measurement of the absolute specific radioactivity of the compounds concerned^{18–22} (which requires accurate chemical determinations of the concentrations of these rare compounds); or (2) a detailed analysis of their initial rates of accumulation in response to the stimulus¹⁹.

Early kinetics of FMLP-driven changes

To define the metabolic origins of the PtdIns(3,4)P₂ and PtdIns(3)P generated in human neutrophils, the rates of change in the levels of all of the ^{32}P -inositol phospholipids were determined 0–2 min after the addition of a maximally effective dose of FMLP (Fig. 3). The curves describing the appearance of the family of '3-phosphorylated' inositol lipids show that the production of PtdIns(3)P and PtdIns(3,4)P₂ lags behind that of PtdIns(3,4,5)P₃. The simplest explanation is that PtdIns(3)P and PtdIns(3,4)P₂ are formed by the stepwise dephosphorylation of PtdIns(3,4,5)P₃.

Metabolism in broken cell systems

Although broken cell assays using exogenous phospholipids as substrates should not be used as the only source of data when defining metabolic networks, they can confirm the existence of enzymatic components of a pathway in cells. We sought evidence for the metabolic fate of PtdIns(3,4,5)P₃ and PtdIns(3,4)P₂ in human neutrophils by introducing 3-[^{32}P]PtdIns(3,4,5)P₃ and 3-[^{32}P]PtdIns(3,4)P₂ into various broken cell systems under conditions as physiological as possible (Fig. 4). The ^{32}P lipids were incorporated into unilamellar vesicles containing a balance of phospholipids imitating that found in the inner leaflet of the

plasma membrane, and assays were conducted in an ionic environment like that of the cytosol.

When 3-[^{32}P]PtdIns(3,4,5) P_3 was incubated with whole lysates under these conditions, it was rapidly degraded (2.5% per min at a protein concentration of $100\text{ }\mu\text{g ml}^{-1}$ compared with 1.65% per min for PtdIns(4,5) P_2 in the same assays). During these assays, 89% of ^{32}P products were recovered in a PtdIns P_2 which was identified as PtdIns(3,4) P_2 (Fig. 4). The remaining 11% of the recovered ^{32}P -products were in P_i which could mean a maximum of 11% of the metabolism of PtdIns(3,4,5) P_3 was yielding PtdIns(4,5) P_2 . But as these lysates can release $^{32}\text{P}_i$ from 3-[^{32}P]PtdIns(3,4) P_2 (see below, and L.R.S., unpublished data) it is likely that this is a quantitatively minor degradative route. The highly active PtdIns(3,4,5) P_3 -5-phosphomonoesterase(s) responsible for most of the metabolism of PtdIns(3,4,5) P_3 was membrane-associated. Typically, 85–90% of the activity recovered from lysates was particulate and could be differenti-

ated from the PtdIns(4,5) P_2 -5-phosphomonoesterase activity in the same membrane preparations by resistance to 10 mM EDTA. The activity (or activities) that released $^{32}\text{P}_i$ from 3-[^{32}P]PtdIns(3,4,5) P_3 was predominantly soluble (92–95% of the activity from fractionated lysates) and was apparently activated by chelation of divalent cations in the assay buffer by EDTA (like a soluble PtdIns(3)P-phosphatase²³).

We also used 3-[^{32}P]PtdIns(3,4) P_2 as a substrate for neutrophil lysates. Its rate of metabolism was lower (0.22% per min, at a protein concentration of $100\text{ }\mu\text{g ml}^{-1}$) than that of PtdIns(3,4,5) P_3 and PtdIns(4,5) P_2 . Of the total PtdIns(3,4) P_2 phosphatase activity recovered from fractionated lysates, 75–80% was soluble, and in assays conducted with these fractions, 30–40% of the ^{32}P liberated from the substrate was contained in PtdIns(3)P (see legend to Fig. 4). The activity causing the release of $^{32}\text{P}_i$ from 3-[^{32}P]PtdIns(3,4) P_2 was soluble and activated by EDTA, as are the analogous activities releasing $^{32}\text{P}_i$

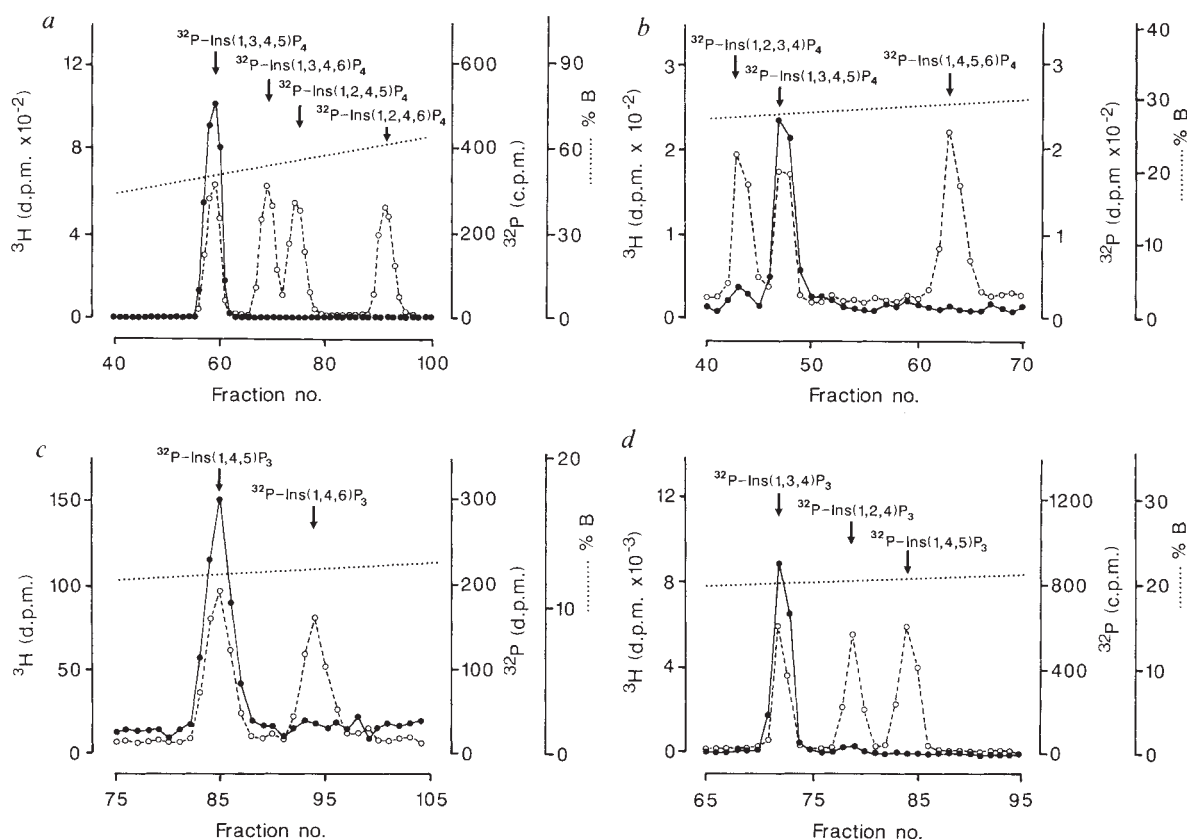


FIG. 2 HPLC analysis of the inositol phosphates derived from peaks 12 and 3 in Fig. 1. $\circ$, ^3H ; $\bullet$, ^{32}P . a, The ^3H -metabolite generated by deglyceration of ^3H -GroPIns P_3 (peak 3 in Fig. 1) was mixed with ^{32}P -labelled Ins(1,3,4,5) P_4 , Ins(1,3,4,6) P_4 , Ins(1,2,4,5) P_4 and Ins(1,2,4,6) P_4 , applied to a Partisphere 5-SAX volume and eluted at 1 ml min^{-1} with a gradient of $2\text{ M NaH}_2\text{PO}_4$ (pH 4.4 with NaOH) (%B). Fractions were collected every 34 s. b, An aliquot of the same preparation of deglycerated ^3H -GroPIns P_3 resolved in a was mixed with ^{32}P -labelled Ins(1,2,3,4) P_4 , Ins(1,3,4,5) P_4 and Ins(1,4,5,6) P_4 applied to a Partisphere 5-WAX column and eluted at 1 ml min^{-1} with a gradient of $0.5\text{ M (NH}_4)_2\text{HPO}_4$ (pH 3.2 with H_3PO_4) (%B). Fractions were collected every 45 s. c, The ^3H -Ins P_3 produced by the sequential deglyceration and dephosphorylation (with human erythrocyte plasma membranes in the presence of EDTA) of ^3H -GroPIns P_3 was mixed with ^{32}P -labelled Ins(1,4,5) P_3 (some ^{32}P -Ins(1,4,5) P_3 was already present in the samples) and Ins(1,4,6) P_3 (see below), applied to a Partisphere 5-SAX HPLC column and eluted at 1 ml min^{-1} with a gradient based on $2\text{ M NaH}_2\text{PO}_4$ (pH 4.4 with NaOH) (%B). Fractions were collected every 35 s. d, The ^3H -Ins P_3 derived from the deglyceration of the agonist-sensitive ^3H -GroPIns P_2 (peak 1 in Fig. 1) was mixed with ^{32}P -labelled Ins(1,3,4) P_3 , Ins(1,2,4) P_3 and Ins(1,4,5) P_3 and applied to a Partisphere 5-WAX HPLC column and eluted at 1 ml min^{-1} with a

gradient based on $0.5\text{ M (NH}_4)_2\text{HPO}_4$ (pH 3.8 with H_3PO_4) (%B). Fractions were collected every 40 s.

METHODS. ^{32}P -labelled Ins(1,4,5) P_3 , Ins(1,3,4) P_3 , Ins(1,3,4,5) P_4 , Ins(1,3,4,6) P_4 and Ins(3,4,5,6) P_4 were prepared as described^{18,31,33,34}. ^{32}P -Ins(1,2,3,4) P_4 was prepared as the first ^{32}P -Ins P_4 derived from an extract of ^{32}P -labelled mung beans eluted from a Partisphere 5-SAX HPLC column (it emerges just before internal ^3H -Ins(1,3,4,5) P_4 when the column is eluted with $1\text{ M NaH}_2\text{PO}_4$, pH 3.8). ^{32}P -Ins(1,2,4,5) P_4 , ^{32}P -Ins(1,2,4,6) P_4 and ^{32}P -Ins(1,2,4) P_3 were prepared by treating ^{32}P -Ins(1,3,4,5) P_4 , ^{32}P -Ins(1,3,4,6) P_4 and ^{32}P -Ins(1,3,4) P_3 with 1 M HCl (20 min at 80°C) then purifying the products by HPLC; their identities were established (1) by treating parallel ^3H -Ins-labelled extracts with periodate then identifying the polyols generated after reduction and dephosphorylation^{16,35} and (2) by comparison of their chromatographic properties with ^3H -labelled standards. ^3H -Ins(1,2,4) P_3 yields ^3H -altritol and eluted after Ins(1,3,4) P_3 and before Ins(1,4,5) P_3 . ^3H -Ins(1,2,4,5) P_4 yielded ^3H -Ins and was the earlier eluting of the two products of the acid migration. ^3H -Ins(1,2,4,6) P_4 was identified as described²³. ^{32}P -Ins(1,4,6) P_3 was prepared by incubating ^{32}P -Ins(1,3,4,6) P_4 with human erythrocyte plasma membranes (with 5 mM EDTA , other conditions precisely as described in the legend to Fig. 1); ^{32}P -Ins(1,4,6) P_3 constituted 20–30% of the products.

from 3-[^{32}P]-PtdIns(3,4,5) P_3 and 3-[^{32}P]-PtdIns3P (ref. 23, and L.R.S. unpublished data).

Structures and synthesis

It is probable that the structures we have unambiguously derived for the inositol phospholipids found in FMLP-stimulated human neutrophils (and previously for PtdIns(3)P in human astrocytoma cells¹⁷) will generally be accepted to account for chromatographically similar compounds already identified in cultured cells and other tissues^{3-6,24,25}. The metabolic source of these compounds in the intact cell is a more difficult issue. Two different mechanisms are currently proposed to account for the agonist-driven rises in the 3-phosphorylated inositol lipids that occur in PDGF-stimulated smooth muscle^{11,13} and thrombin-stimulated platelets^{15,26} (Fig. 5b and c, respectively). Our data are partially consistent with the former model but are not consistent with the latter. The ^{32}P -labelling data suggest that PtdIns(3,4,5) P_3 synthesized in response to FMLP was formed by the 3-phosphorylation of PtdIns(4,5) P_2 . But the PtdIns(3,4) P_2 and PtdIns(3)P produced concurrently could have been derived either from the parallel phosphorylation of PtdIns(4)P and/or PtdIns or from the dephosphorylation of PtdIns(3,4,5) P_3 or PtdIns(3,4) P_2 . We suggest that the PtdIns(3,4) P_2 and PtdIns(3)P are derived from the sequential dephosphorylation of PtdIns(3,4,5) P_3 because of (1) the kinetics of the appearance of PtdIns(3)P and PtdIns(3,4) P_2 in FMLP-

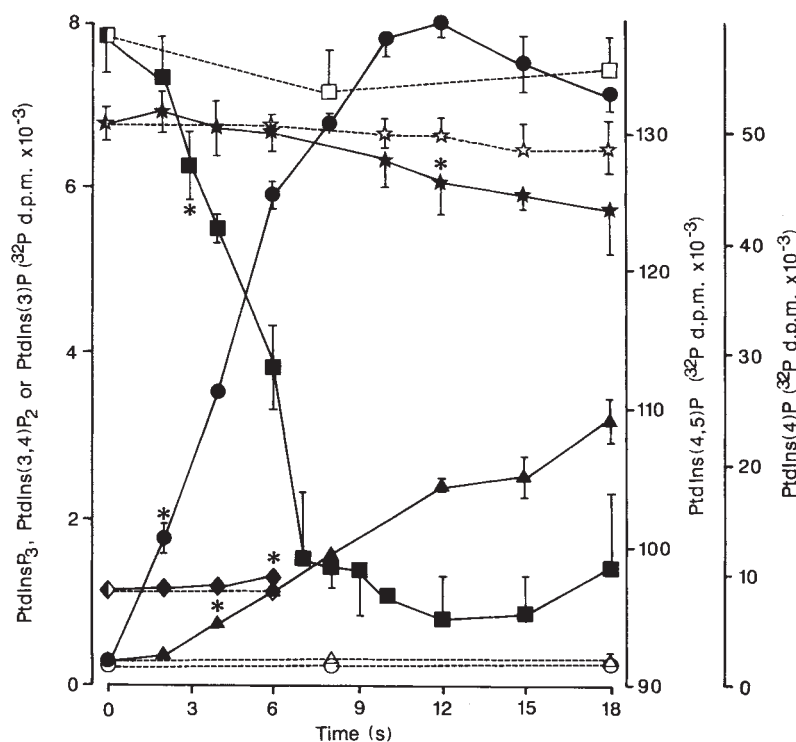
stimulated neutrophils (Fig. 3) and (2) the activities in broken-neutrophil preparations that can catalyse the appropriate dephosphorylation reactions.

This suggestion differs from the mechanism proposed by Carpenter and Cantley¹⁴ (Fig. 5b). Their map was based on experimental evidence that showed that, when the PtdIns-3-OH kinase (translocated and tyrosine phosphorylated in response to PDGF) is assayed *in vitro*, it readily phosphorylates PtdIns and PtdIns(4)P^{3,4,13,14,25}. We cannot discount the possibility that direct 3-phosphorylation of PtdIns and PtdIns(4)P occurs in stimulated neutrophils, but the most parsimonious explanation of our data does not require this to occur. Indeed, to explain the formation of PtdIns(3)P and PtdIns(3,4) P_2 by direct phosphorylation of PtdIns and PtdIns(4)P, we would also have to propose that the substrate specificity and/or substrate supply of the agonist-activated 3-kinase is different from that *in vitro*, and moreover that it changes during the first few seconds in its activated state.

The history of the assay of phosphoinositidase C provides a direct precedent for the idea of a discrepancy between the substrate specificity of phospholipid-metabolizing enzymes *in vitro* and *in vivo*. Phosphoinositidases C were originally purified and classified as PtdIns-selective phospholipases C on the basis of their preference for PtdIns *in vitro*. But it was only after these same enzymes were shown to display a marked preference for the polyphosphoinositides *in vivo*^{27,28} that it was appreciated

FIG. 3 Rates of change of levels of inositol phospholipids in control (open symbols) and FMLP-stimulated (filled symbols) human neutrophils. PtdIns P_3 (circles), PtdIns(3,4) P_2 (triangles), PtdIns(3)P (diamonds), PtdIns(4,5) P_2 (squares) and PtdIns(4)P (asterisks).

METHODS. Human neutrophils were prepared as described (ref. 36; except that the high-density Percoll layer was 1.099 g ml^{-1} and the physiological salt solutions were all as defined in ref. 2) and incubated with ^{32}P (5×10^7 cells ml^{-1} $0.4\text{--}2.4\text{ mCi ml}^{-1}$ ^{32}P ; PBS-13, Amersham) with no added calcium and with 2 mg ml^{-1} BSA. After 70 min the cells were washed twice with fresh medium, 200–300- μl aliquots were equilibrated (under conditions identical to those in which they were labelled) for 10 min, then mixed with pre-warmed aliquots of medium (as above) either with or without FMLP (final concentration in the cell suspension was $1\text{ }\mu\text{M}$), quenched and finally a phospholipid extract prepared. The phospholipid extraction protocol was as in refs 1 and 2, except that first, the methanol-HCl solution with which the phospholipid extracts were finally washed also contained 50 mM EDTA, 1 mM cyclohexane-1,2-diamine tetra-acetic acid (CDTA), 1 mM H_3PO_4 and 1 mM myo-inositol; and second, all of the surfaces with which the samples came into contact were polypropylene. Several other extraction protocols were tested^{37,38} and none recovered more ^{32}P -PtdIns P_3 from a given suspension of cells and some (as in ref. 37) substantially less. To quantify all ^{32}P -phospholipids (except the PtdInsPs; see below) the dry lipid films were deacylated³⁹ (a ratio of 500 μl deacylating reagent, prepared from mono-methylamine gas, to lipid derived from 1.5×10^7 cells was used throughout) and the water-soluble compounds generated were mixed with standards (all prepared either from ^3H -Ins-labelled cells or from commercially available standards as described^{17,40}) and resolved on an anion-exchange HPLC column (Fig. 1). The ^{32}P compounds resolved were identified by reference to internal ^3H -standards and/or the structural analysis shown in the legends to Figs 1 and 2. If neutrophil-derived phospholipid extracts are deacylated³⁹ directly and analysed by HPLC, then although all other regions of the chromatographic profile elute perfectly reproducibly (both with respect to internal ^3H -labelled standards and the endogenous ^{32}P -metabolites), the GroPInsP region eluted at variable times and often as two peaks; when this occurred internally, demonstrably pure ^3H -Ins-labelled standards also gave two peaks. This resulted from material derived from the neutrophils that eluted in this region of the HPLC profile, but could be eliminated by purifying the PtdInsPs by TLC before deacylation and HPLC (TLC plates were developed as described¹ except that they were sprayed with 1% potassium oxalate). All PtdIns(4)P and PtdIns(3)P analyses were performed with lipids that had been TLC-



purified before being deacylated and finally resolved by anion-exchange HPLC as described above. To quantify the recovery of ^{32}P -PtdInsPs through this process, all extracts were mixed with 10,000 d.p.m. of ^3H -PtdIns(4)P (Amersham; recoveries ranged from 82–96%). The data shown are derived from eight experiments and are normalized by reference to the ^{32}P -Pi concentration in the final cell suspensions during the labelling process. Individual points are means $\pm$ s.e. m. ($n=3\text{--}10$, error bars that fell within the symbols are omitted for clarity). Asterisks indicate the first times at which particular compounds became different from controls ($P<0.05$). Data for ^{32}P -PtdIns(3,4) P_2 , ^{32}P -PtdIns(3,4,5) P_3 and ^{32}P -PtdIns(4,5) P_2 contained within those samples that were initially separated on TLC were not included, although they showed exactly the same pattern of change. All of the data points shown are based on measurements of a minimum of 200 d.p.m. ^{32}P above background.

METHODS. Assay of 3-[^{32}P]-PtdIns(3,4,5) P_3 , 3-[^{32}P]-PtdIns(3,4) P_2 and ^3H -PtdIns(4,5) P_2 metabolism was as follows. The ^{32}P -lipids were prepared and purified on an amino-propyl HPLC column using 1 M ammonium acetate in chloroform: methanol: H_2O (10:9:1 (v/v/v); buffer B, see ref. 41). The lipids were stored in buffer B under liquid nitrogen. In some experiments ^3H -PtdIns(4,5) P_2 was mixed with the above substrates before adding them to a solution of phospholipids in chloroform (containing by mol per cent, 26.5% PtdIns, 26.5% PtdE, 12% PtdCho, 25.1% PtdSer (Sigma), 2.8% sphingomyelin (Sigma), 2.3% PtdIns(4,5) P_2 , prepared as described^{42,43}). Unilamellar vesicles were prepared from solutions of phospholipids in chloroform by evaporating the solvent under vacuum then bath-sonicating the lipid films into 20 mM HEPES, 1 mM EGTA, 0.1 mM EDTA (pH 7.2, 25 $^\circ\text{C}$)³². Assays of 50 μl contained between 15,000 d.p.m. of ^{32}P -lipid (some assays also contained 30,000 d.p.m. of ^3H -PtdIns(4,5) P_2), 0.2 mg ml^{-1} unilamellar phospholipid vesicles, 1 mg ml^{-1} BSA, 80 mM KCl, 5 mM MgCl_2 , 1 mM EGTA, 20 mM HEPES (pH 7.2, 37 $^\circ\text{C}$) and aliquots of neutrophil lysates, soluble or particulate fractions⁴⁴. Assays were run for 0.5–4.0 min (such that less than 5% of the substrate had been hydrolysed) quenched and phospholipids extracted and either deacylated or resolved by TLC (see above and the legend to Fig. 3). Water-soluble ^3H and/or ^{32}P compounds released during these assays were collected from the aqueous phases of the phospholipid extractions, neutralized and resolved by anion-exchange HPLC. The only ^{32}P -labelled product detected was $^{32}\text{P}_i$, no water-soluble ^3H -metabolites

were released during any of the assays. The products of the metabolism of PtdIns(3,4,5)P₃ and PtdIns(3,4)P₂ were identified as follows. The ³²P-product produced from 3-[³²P]-PtdIns(3,4,5)P₃ by neutrophil particulate fractions (lane 2) ran at the trailing edge of the PtdIns(4,5)P₂ used as an internal carrier and was likely to be a PtdInsP₂ (on the basis of both the substrate from which it is derived and its chromatographic properties compared to standards, and moreover because it had retained its ³²P label, it must contain a 3-phosphate). When samples of the ³²P-PtdInsP₂ were excised from TLC plates, deacylated and the water-soluble products resolved by HPLC (see legends to Figs 1 and 3) the only detectable ³²P-labelled metabolite eluted just prior to internal ³H-GroPIns(4,5)P₂ and at precisely the time expected for GroPIns(3,4)P₂ (data not shown). The ³²P-labelled lipid product derived from the dephosphorylation of 3-[³²P]-PtdIns(3,4)P₂ by neutrophil cytosolic fractions was identified as PtdIns(3)P on the following grounds: the structure and position of the ³²P-label in the substrate from which it was derived, its chromatographic mobility compared to known compounds during TLC (data not shown), and finally, when deacylated the water-soluble head group co-migrated with internal ³H-GroPIns(3)P during anion-exchange HPLC (the ³H standard was prepared from ³H-Ins-labelled astrocytoma cells¹⁷; data not shown).

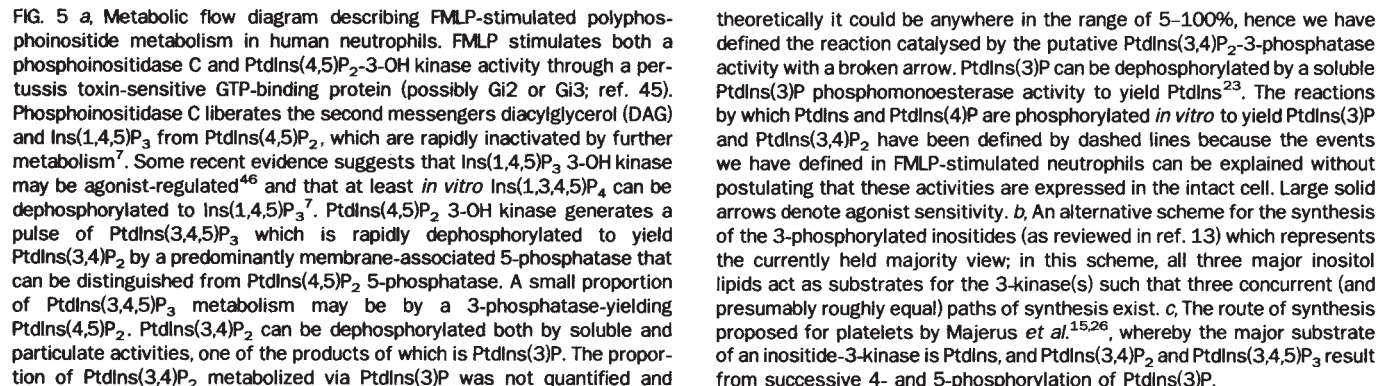


FIG. 5. A Metabolic flow diagram describing FMLP-stimulated polyphosphoinositide metabolism in human neutrophils. FMLP stimulates both a phosphoinositidase C and PtdIns(4,5) P_2 -3-OH kinase activity through a pertussis toxin-sensitive GTP-binding protein (possibly Gi2 or Gi3; ref. 45). Phosphoinositidase C liberates the second messengers diacylglycerol (DAG) and Ins(1,4,5) P_3 from PtdIns(4,5) P_2 , which are rapidly inactivated by further metabolism⁷. Some recent evidence suggests that Ins(1,4,5) P_3 -3-OH kinase may be agonist-regulated⁴⁶ and that at least *in vitro* Ins(1,3,4,5) P_4 can be dephosphorylated to Ins(1,4,5) P_3 ⁷. PtdIns(4,5) P_2 -3-OH kinase generates a pulse of PtdIns(3,4,5) P_3 which is rapidly dephosphorylated to yield PtdIns(3,4) P_2 by a predominantly membrane-associated 5-phosphatase that can be distinguished from PtdIns(4,5) P_2 5-phosphatase. A small proportion of PtdIns(3,4,5) P_3 metabolism may be by a 3-phosphatase-yielding PtdIns(4,5) P_2 . PtdIns(3,4) P_2 can be dephosphorylated both by soluble and particulate activities, one of the products of which is PtdIns(3)P. The proportion of PtdIns(3,4) P_2 metabolized via PtdIns(3)P was not quantified and

theoretically it could be anywhere in the range of 5–100%, hence we have defined the reaction catalysed by the putative PtdIns(3,4)P₂-3-phosphatase activity with a broken arrow. PtdIns(3)P can be dephosphorylated by a soluble PtdIns(3)P phosphomonoesterase activity to yield PtdIns²³. The reactions by which PtdIns and PtdIns(4)P are phosphorylated *in vitro* to yield PtdIns(3)P and PtdIns(3,4)P₂ have been defined by dashed lines because the events we have defined in FMLP-stimulated neutrophils can be explained without postulating that these activities are expressed in the intact cell. Large solid arrows denote agonist sensitivity. *b*, An alternative scheme for the synthesis of the 3-phosphorylated inositides (as reviewed in ref. 13) which represents the currently held majority view; in this scheme, all three major inositol lipids act as substrates for the 3-kinase(s) such that three concurrent (and presumably roughly equal) paths of synthesis exist. *c*, The route of synthesis proposed for platelets by Majerus *et al.*^{15,26}, whereby the major substrate of an inositide-3-kinase is PtdIns, and PtdIns(3,4)P₂ and PtdIns(3,4,5)P₃ result from successive 4- and 5-phosphorylation of PtdIns(3)P.

that the specificity of phosphoinositidases C *in vitro* are critically dependent on the assay conditions²⁹.

The model proposed to account for PtdIns(3,4,5)P₃ production in thrombin-stimulated platelets^{15,26} (Fig. 5c) is incompatible with our data. This could be a result of a cell- or agonist-related difference. The only published information describing the kinetics of this response in the intact thrombin-stimulated platelet clearly shows that PtdInsP₃ levels rose and began to fall before PtdIns(3,4)P₂ concentrations significantly increased⁵. This result would be difficult to reconcile with the synthesis of PtdInsP₃ from PtdIns(3,4)P₂. Clearly, the issue of whether tissues actually differ, or whether experimental protocols account for the difference, remains to be resolved.

Dynamics of PtdIns(3,4,5)P₃ response

The relative flux of material through this pathway and through phosphoinositidase C in stimulated neutrophils can be calculated. We have to assume (1) that the initial rate of agonist-driven depletion of PtdIns(4,5)P₂ represents the sum of the fluxes through phosphoinositidase C and PtdIns(4,5)P₂-3-OH kinase, and (2) that the amounts of radioactivity contained in these

lipids in the experiments defined in Fig. 3 reflect their concentrations (allowing for the different complements of monoester phosphates). Assuming these, then 15–20% of the PtdIns(4,5)P₂ that is lost during the first 6 s of stimulation with a maximally effective dose of FMLP flows through PtdIns(4,5)P₂-3-OH kinase. The concentration of PtdIns(4,5)P₂ in resting neutrophils is 50 nmol per ml packed cells (L.R.S., unpublished data). Hence the maximal initial flux through PtdIns(4,5)P₂-3-OH kinase is about 0.34 nmol s⁻¹ ml⁻¹ packed cells and the concentration of PtdIns(3,4,5)P₃ rises from a basal value of 50 pmol ml⁻¹ to a peak value of 2 nmol ml⁻¹. These values suggest intracellular concentrations of PtdInsP₃ of the order of 5 μ M (unstimulated) and 200 μ M (stimulated), provided that all of this PtdIns(3,4,5)P₃ is present in the inner leaflet of the plasma membrane³⁰.

The ability of agonists to produce such a localized pulse of a molecule with the structural characteristics of PtdIns(3,4,5)P₃ clearly supports claims that this system may represent a new signalling pathway^{13–15} and hence establishes PtdIns(3,4,5)P₃-3-OH kinase as an enzyme of enormous potential interest. $\square$

Received 6 December 1990; accepted 22 March 1991.

1. Traynor-Kaplan, A. E., Harris, A. L., Thompson, B., Taylor, P. & Sklar, L. A. *Nature* **334**, 353–356 (1988).
2. Traynor-Kaplan, A. E. *et al.* *J. biol. Chem.* **264**, 15668–15673 (1989).
3. Auger, K. R., Suranian, L. A., Soltoff, S. P., Libby, P. & Cantley, L. *Cell* **57**, 167–175 (1989).
4. Ruderman, N. B., Kapeller, R., White, M. F. & Cantley, L. *Proc. natn. Acad. Sci. U.S.A.* **87**, 1411–1415 (1990).
5. Kucera, G. L. & Rittenhouse, S. E. *J. biol. Chem.* **265**, 5245–5248 (1990).
6. Ascoli, M. & Pignataro, P. P. *J. biol. Chem.* **265**, 1718–1723 (1990).
7. Berridge, M. J. & Irvine, R. F. *Nature* **341**, 197–205 (1989).
8. Whitman, M., Kaplan, D. R., Schaffhausen, B., Cantley, L. & Roberts, T. M. *Nature* **315**, 239–241 (1985).
9. Whitman, M., Kaplan, D. R., Roberts, T. & Cantley, L. *Biochem. J.* **247**, 165–174 (1987).
10. Endomann, G., Yonezawa, K. & Roth, R. A. *J. biol. Chem.* **265**, 396–400 (1990).
11. Kaplan, D. R. *et al.* *Cell* **50**, 1021–1029 (1987).
12. Courtneidge, S. A. & Heber, A. *Cell* **50**, 1031–1037 (1987).
13. Whitman, M. & Cantley, L. *Biochim. biophys. Acta* **948**, 327–344 (1988).
14. Carpenter, C. L. & Cantley, L. C. *Biochemistry* **29**, 11147–11156 (1990).
15. Majerus, P. W. *et al.* *Cell* **63**, 459–465 (1990).
16. Tomlinson, R. V. & Ballou, C. E. *J. biol. Chem.* **236**, 1902–1906 (1961).
17. Stephens, L. R., Hawkins, P. T. & Downes, C. P. *Biochem. J.* **259**, 267–276 (1989).
18. Stephens, L. R. & Downes, C. P. *Biochem. J.* **265**, 435–452 (1990).
19. Stephens, L. R., Berrie, C. P. & Irvine, R. F. *Biochem. J.* **269**, 65–72 (1990).
20. Brockerhoff, H. & Ballou, C. E. *J. biol. Chem.* **237**, 1764–1766 (1962).
21. King, C. E., Hawkins, P. T., Stephens, L. R. & Michell, R. H. *Biochem. J.* **259**, 893–898 (1989).
22. King, C. E., Stephens, L. R., Hawkins, P. T., Guy, C. R. & Michell, R. H. *Biochem. J.* **244**, 209–217 (1988).
23. Lips, D. L. & Majerus, P. W. *J. biol. Chem.* **264**, 19911–19915 (1989).
24. Vadnol, R. E. & Parthasarathy, R. *Biochem. biophys. Res. Comm.* **163**, 995–1001 (1989).
25. Morgan, S., Smith, A. D. & Parker, P. J. *Eur. J. Biochem.* **191**, 761–767 (1990).

26. Cunningham, T. W. *et al.* *J. biol. Chem.* **265**, 21676–21683 (1990).
27. Allen, D. & Michell, R. H. *Biochim. biophys. Acta* **508**, 277–286 (1978).
28. Downes, C. P. & Michell, R. H. *Biochem. J.* **198**, 133–140 (1981).
29. Irvine, R. F., Letcher, A. J. & Dawson, R. M. C. *Biochem. J.* **218**, 177–185 (1984).
30. Downes, C. P. & Michell, R. H. *Cell Calcium* **3**, 467–502 (1982).
31. Letcher, A. J., Stephens, L. R. & Irvine, R. F. (1990) in *Methods in Inositol Research* (ed. Irvine, R. F.) 31–37 (Raven, New York, 1990).
32. Auger, K. R., Suranian, L. A. & Cantley, L. C. (1990) in *Methods in Inositol Research* (ed. Irvine, R. F.) 159–166 (Raven, New York, 1990).
33. Stephens, L. R. (1990) in *Methods in Inositol Research* (ed. Irvine, R. F.) 9–30 (Raven, New York, 1990).
34. Stephens, L. R., Hawkins, P. T. & Downes, C. P. *Biochem. J.* **262**, 727–737 (1989).
35. Stephens, L. R. & Irvine, R. F. *Nature* **364**, 580–583 (1990).
36. Merritt, J. E., Jacob, R. & Hallam, T. J. *J. biol. Chem.* **264**, 1522–1527 (1989).
37. Creba, J. A. *et al.* *Biochem. J.* **212**, 733–747 (1983).
38. Bligh, E. G. & Dyer, W. J. *Can. J. Biochem. Physiol.* **37**, 911–917 (1959).
39. Clarke, N. G. & Dawson, R. M. C. *Biochem. J.* **165**, 301–306 (1981).
40. Stephens, L. R., Hawkins, P. T., Barker, C. J. & Downes, C. P. *Biochem. J.* **253**, 721–733 (1988).
41. Low, M. G., Carroll, R. C. & Cox, A. C. *Biochem. J.* **237**, 139–145 (1986).
42. Irvine, R. F., Hemington, N. & Dawson, R. M. C. *Biochem. J.* **176**, 475–484 (1978).
43. Palmer, F. B. St C. *J. Lipid Res.* **22**, 1296–1300 (1981).
44. Bennett, J. P., Cockcroft, S., Caswell, A. H. & Gomperts, B. D. *Biochem. J.* **208**, 801–807 (1982).
45. Milligan, G. *Biochem. J.* **255**, 1–13 (1988).
46. Sim, S. S., Kim, J. W. & Rhee, S. G. *J. biol. Chem.* **265**, 10367–10372 (1990).
47. Imboden, J. B. & Pattison, G. *J. clin. Invest.* **79**, 1538–1541 (1987).
48. Takazawa, K., Passareiro, H., Dumont, J. E. & Erneux, C. *Biochem. J.* **261**, 483–488 (1989).
49. Inhorn, R. C., Bansal, V. S. & Majerus, P. W. *Proc. natn. Acad. Sci. U.S.A.* **84**, 2170–2174 (1987).
50. Gee, N. S., Reid, G. G., Jackson, R. G., Barnaby, R. J. & Ragan, C. I. *Biochem. J.* **253**, 777–782 (1988).

ACKNOWLEDGEMENTS. We thank P. Hawkins, T. Jackson and C. MacPhee for helpful discussions and R. Smith for taking blood samples. L.R.S. is a Babraham Research Fellow.

LETTERS TO NATURE

Orbital evolution of low-mass X-ray binaries due to radiation driven mass transfer

Marco Tavani

Institute of Geophysics and Planetary Physics,
Lawrence Livermore National Laboratory, California 94550, USA;
and Department of Astronomy, University of California at Berkeley,
California 94720, USA

A LOW-MASS X-ray binary (LMXB) consists of a compact star, probably a neutron star, accreting mass from a low-mass ($\leq 1 M_{\odot}$) companion via an accretion disk. Of ~ 100 known LMXBs in the Galaxy, only four have stable enough X-ray modulations to have allowed the reliable determination of orbital period changes. For these four LMXBs, all of which have $P_{\text{orb}} \leq 5.6$ h, the measured values^{1–4} of $\dot{P}_{\text{orb}}/P_{\text{orb}}$ disagree markedly with what would be expected for orbital evolution driven by angular momentum loss due to gravitational radiation, possibly supplemented by magnetic

braking; the empirically derived timescale for orbital evolution is ~ 100 times less than expected. On the assumption that the observed period changes are secular, and not due to some longer-term periodic change, I argue here that the observed behaviour of LMXBs can be explained as the result of mass loss from the companion star caused by irradiation of the secondary star and accretion disk by the primary⁵. The typical lifetime of a radiation-driven LMXB is expected to be $\sim 10^6$ – 10^7 yr. This reduced evolutionary timescale can resolve the statistical discrepancy between the number of binary millisecond pulsars and of their presumed LMXB progenitors if about half of all the LMXBs are radiation-driven^{5–7}.

The mechanism driving LMXB mass transfer depends greatly upon the nature of the companion star. In models of mass transfer neglecting the effect of radiation from the primary, it is assumed that the companion always fills its Roche lobe. Evolutionary models that neglect the effect of radiation have guided studies of the X-ray emission properties of several individual LMXBs. In the orbital period range of interest here, the mass transfer is driven by the loss of orbital angular momentum due to gravitational radiation (GR)^{8,9}, possibly supplemented by the 'magnetic braking' effect (MB) if the

sequence companion has a sufficiently large magnetic field and intrinsic mass loss^{10–11}. I will refer to this evolutionary model as the ‘standard’ model. The typical timescale for GR evolution is

$$\tau_{\text{GR}} \approx (5 \times 10^9 \text{ yr}) a_{11}^{-4} (M/M_\odot)^{-2} (m/M_\odot)^{-1} (1+m/M)^{-1}$$

where $a_{11} = a/(10^{11} \text{ cm})$ and a is the orbital radius, and M and m are the masses of the primary and companion star respectively. Magnetic braking may be important in the case of companion stars with radiative cores, in other words main sequence^{10,11} and possibly sub-giant stars¹² that lose mass to a wind which is magnetically coupled to the star. Here the relevant timescale¹⁰ is

$$\tau_{\text{MB}} \approx (1.8 \times 10^9 \text{ yr}) (B_s/200 \text{ G})^{-4/3} (R/R_\odot)^{-8/3} \times (\dot{m}_w/10^{-10} M_\odot \text{ yr}^{-1})^{-1/3} (M/1.4 M_\odot) (m/0.6 M_\odot)$$

where B_s is the surface magnetic field of the companion, R the companion’s radius, $R_\odot$ the solar radius, and $\dot{m}_w$ the intrinsic mass loss rate from the companion. The evolutionary timescale, τ_{evol} , predicted by the standard model from the characteristics of the LMXBs in Table 1 is $10^9 \text{ yr} \leq \tau_{\text{evol}} \leq 10^8 \text{ yr}$.

Measuring $\dot{P}_{\text{orb}}$ for a LMXB requires a stable fiducial point in the X-ray light curve. Such a measurement is often difficult because most LMXBs have no eclipses or pulsations. Furthermore, LMXBs have been observed for $\sim 20 \text{ yr}$, and the lower limit for the measurable $\dot{P}_{\text{orb}}/P_{\text{orb}}$ is $\sim 10^{-7} \text{ yr}^{-1}$. Therefore, no LMXB with measurable $\dot{P}_{\text{orb}}/P_{\text{orb}}$ is expected if the standard model applies to all LMXBs. But reliable non-zero measurements of the time derivative of the orbital period have recently become available^{1–4} for four LMXBs, whose properties are given in Table 1. The measured $\dot{P}_{\text{orb}}$ s, assumed here to represent secular changes of P_{orb} of these sources, disagree with the standard model. This is particularly the case for 4U1820–30, with $P_{\text{orb}} \sim 11 \text{ min}$ ¹³, for which $\dot{P}_{\text{orb}}/P_{\text{orb}}$ has the opposite sign to that expected if it underwent binary evolution with a degenerate, Roche-lobe-filling companion (such as GR evolution¹⁴). In the standard model, binary evolution is given by $\dot{P}_{\text{orb}}/P_{\text{orb}} = [(1-3n)/2] \dot{m}/m$, where the stellar index n is defined by the relation $d \ln R = n d \ln m$, with R the stellar radius. For a Roche-lobe-filling degenerate companion star fitting the orbital parameters of 4U 1820–30, $n = -\frac{1}{3}$, and therefore $\dot{P}_{\text{orb}}/P_{\text{orb}}$ should be positive, contrary to observations.

Special physical effects influencing the LMXBs in question might explain the observed $\dot{P}_{\text{orb}}/P_{\text{orb}}$ in the framework of the standard evolutionary model, but these would need to be different for different sources. They could involve, for example, a fast expansion of the secondary (possibly caused by irradiation⁵), an unusually large surface magnetic field of the companion (as in EXO0748–676), or third-body influence in a triple system or in a globular cluster (as in 4U1820–30; ref. 4).

Here I propose an interpretation of the measured $\dot{P}_{\text{orb}}/P_{\text{orb}}$ in terms of the radiation-driven (RD) model of LMXB evolution^{5,15,16} which could be applicable to a relatively large number of LMXBs⁷. The RD model was applied to Cyg X-3 (ref. 17) to explain what was then considered an exceptionally large $\dot{P}_{\text{orb}}/P_{\text{orb}}$. We now know of three more LMXBs with a large absolute value of $\dot{P}_{\text{orb}}/P_{\text{orb}}$, and can assess the more general applicability of RD evolution. Here I aim to show that RD evolution can explain all $\dot{P}_{\text{orb}}/P_{\text{orb}}$ s yet observed and to identify two classes of RD-LMXBs characterized by different values of the mass loss from the binary.

In a typical LMXB the low-mass companion star is illuminated by a relatively large flux of radiation from the primary star with a spectrum that can drive a strong evaporative wind from its outer atmosphere. Irradiating photons of energy ~ 0.3 – 10 keV or of energy $\sim 1 \text{ MeV}$ are the most effective in driving the mass outflow^{5,7,15,16}. When heating in the outer atmosphere of the companion (or disk) is not balanced by bremsstrahlung, H/He recombination and/or metal-line cooling, an evaporative wind is formed¹⁸. The radiation-driven mass loss rate can be

TABLE 1 Characteristics of the four LMXBs with measured $\dot{P}_{\text{orb}}/P_{\text{orb}}$

LMXB	P_{orb} (hr)	$\dot{P}_{\text{orb}}/P_{\text{orb}}$ (yr^{-1})	Timescale (yr)	Ref.
Cyg X-3	4.82	$+(2.20 \pm 0.22) \times 10^{-6}$	5×10^5	1, 28
X1822–371	5.57	$+(3.40 \pm 0.94) \times 10^{-7}$	2.9×10^6	2
EXO0748–676	3.82	$-(2.02 \pm 0.28) \times 10^{-7}$	5×10^6	3
4U1820–30	0.18	$-(1.08 \pm 0.19) \times 10^{-7}$	10^7	4

written^{5,15,16} as $|\dot{m}_{\text{rad}}| = 10^{-17} f \hat{L} \text{ g s}^{-1}$, where the dimensionless quantity f depends on the details of energy deposition, radiation transport and hydrodynamics at the base of the corona, and $\hat{L}$ is the effective luminosity incident on the companion, $\hat{L} = \chi L \Delta \Omega$, with $\Delta \Omega$ the effective solid angle subtended by the companion star to its primary, L the luminosity (in erg s^{-1}) and χ an attenuation factor which takes into account possible absorption screening, scattering or spectral changes in the disk or corona surrounding the compact star. Previous investigations^{18–21} of a radiation-driven mass loss from the first Lagrangian point in the binary system Her X-1 reached the conclusion that the mechanism would not be capable of sustaining a large mass transfer rate given the intensity and flat energy spectrum of Her X-1. But in LMXBs, the energy flux irradiating the companion (~ 10 times that in Her X-1) and quality of the X-ray spectrum (most of the energy is detected in the soft X-ray band) combine to give an evaporative mass outflow rate of^{5,7,16} $\sim 10^{18}$ – 10^{19} g s^{-1} . Both analytical work and numerical calculations⁵ on radiation-driven mass outflows from irradiated companions or outer parts of the accretion disk give a conservative estimate of the range of f for typical LMXBs as $1 \leq f \leq 10$.

A self-sustained (bootstrap) regime of mass transfer may exist if the accretion-powered illumination of the companion causes mass to be transferred to the neutron star at a rate sufficient to sustain that level of illumination^{5,15,16}. If this mechanism of mass transfer occurs in LMXBs, it has two fixed points. For large companion masses, it may sustain a large ‘bootstrapped’ value of the mass loss rate $\dot{m}^*$. Alternatively, if the companion mass falls below a critical mass, m_c , whose value depends on the nature of the companion as well as on the details of illumination and reprocessing⁵, the wind-driven mass loss rate is no longer sufficient to power accretion close to the bootstrapped value, and self-sustained mass transfer ceases completely. The precise value of $\dot{m}^*$ depends on the detailed geometry, hydrodynamics and radiation transport specific to the binary. For the LMXBs of interest here, the expected mass loss is in the range $\sim 10^{18}$ – 10^{19} g s^{-1} with $\chi \sim 0.1$ (ref. 22), and the corresponding X-ray luminosity L_x is in the range $\sim 10^{37}$ – $10^{38} \text{ erg s}^{-1}$ which covers most known LMXBs.

The orbital evolution of an LMXB is determined by angular momentum loss (driven by gravitational radiation or a wind), as well as by mass loss from the companion star. A self-sustained RD mass loss from the companion (or from the outer edge of the disk) markedly changes the standard evolutionary pattern (ref. 23 gives a review of standard orbital evolution). The irradiated companion can transfer mass even if it does not fill its Roche lobe exactly (that is, $\Psi \neq 1$ with $\Psi \equiv R/R_L$, where R_L is the Roche lobe radius of the companion). We obtain

$$\frac{\dot{P}_{\text{orb}}}{P_{\text{orb}}} = -3 \left[1 - (1 - \beta) \tilde{\alpha} (1 + q) - \beta q - \frac{1}{3} (1 - \beta) \frac{q}{1 + q} \right] \times \frac{\dot{m}}{m} + 3 \left[\left(\frac{j}{J} \right)_{\text{GR}} + \left(\frac{j}{J} \right)_{\text{MR}} \right] \quad (1)$$

where $\dot{m}$ is the mass transfer rate, β the fraction of $|\dot{m}|$ which is accreted onto the compact star, $(j/J)_{\text{GR}} + (j/J)_{\text{MR}}$ the fractional change of orbital angular momentum from gravitational radiation and magnetic braking, $q = m/M$ the ratio of the companion mass to primary mass M , and $\tilde{\alpha}$ the specific angular

momentum parameter defined by $\delta J = \tilde{\alpha} \delta m (1 - \beta) a^2 2\pi / P_{\text{orb}}$. The dimensionless parameter $\tilde{\alpha}$ gives the effective angular momentum loss caused by mass loss from the binary. Mass loss from the binary may play an important part in determining the evolution of RD-LMXBs. Complex effects including the geometry, hydrodynamics, final wind velocity, and energy exchange between the outflowing gas and primary star radiation all contribute to the effective value of β . I restrict the analysis here to a few idealized cases which are relevant for the LMXBs of Table 1. For a self-sustained value of the RD mass loss rate $-\dot{m} = |\dot{m}_{\text{rad}}| = (10^{-8} M_{\odot} \text{ yr}^{-1}) \dot{m}_8$, where $\dot{m}_8 = |\dot{m}| / (10^{-8} M_{\odot} \text{ yr}^{-1})$, the first term of the righthand side of equation (1), proportional to $\dot{m}/m$ is dominant, and I will neglect the contribution from GR and MB in the following discussion. For degenerate and main-sequence companions, the time derivative of the orbital period may be either positive or negative depending primarily on the value of β for $\tilde{\alpha}$ of order unity. The associated theoretical lifetime can be made to be $\sim 10^7$ yr or less. Figure 1 gives the computed $\dot{P}_{\text{orb}}/P_{\text{orb}}$ for different values of the mass-loss parameter β and for self-sustained $\dot{m}$ appropriate to the LMXBs of Table 1. The dimensionless parameter $\tilde{\alpha}$ has been fixed to give the specific angular momentum loss at the orbital radius; here

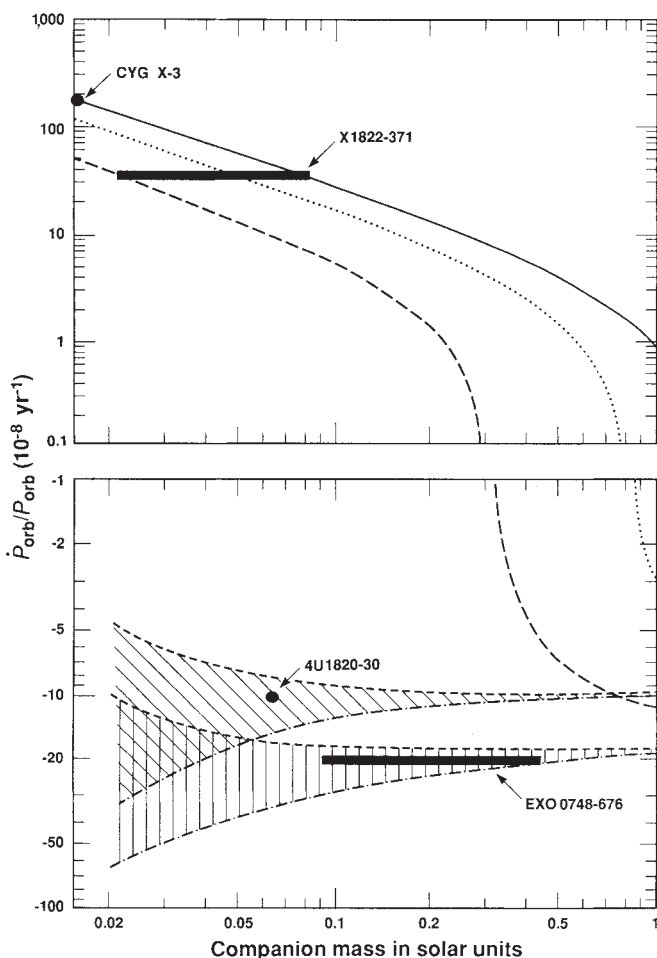


FIG. 1 The quantity $\dot{P}_{\text{orb}}/P_{\text{orb}}$ computed for the RD evolutionary model, as a function of the mass of the companion for $\tilde{\alpha}=1$. The upper part of the diagram has a positive $\dot{P}_{\text{orb}}/P_{\text{orb}}$. The three curves are computed for $\beta=1$ (solid line), $\beta=2/3$ (dotted line) and $\beta=1/3$ (dashed line). The horizontal thick marks, corresponding to the measured $\dot{P}_{\text{orb}}/P_{\text{orb}}$ for Cyg X-3 and X1822-371, give the mass range for the companion for $\dot{m}_8=3$. The lower part of the diagram has a negative $\dot{P}_{\text{orb}}/P_{\text{orb}}$. The two hatched areas give the allowed range for $\dot{P}_{\text{orb}}/P_{\text{orb}}$ from the RD evolutionary model with $\dot{m}_8=10$ and 20, respectively. The upper dashed curve is for $\beta=1/10$ and the lower dot-dashed curve for $1/20$. The horizontal thick marks, corresponding to the measured $\dot{P}_{\text{orb}}/P_{\text{orb}}$ for EXO0748-676 and 4U1820-30, give the mass range for the companion stars.

I use $\tilde{\alpha}=1$. The thick horizontal lines in Fig. 1 which correspond to the measured $\dot{P}_{\text{orb}}/P_{\text{orb}}$ give the mass ranges of the companion obtained from the RD evolutionary model.

Cyg X-3, with $\dot{P}_{\text{orb}}/P_{\text{orb}} = +2 \times 10^{-6} \text{ yr}^{-1}$ (ref. 1, 24, 28) fits the evolutionary scenario for a RD-LMXB containing a very low mass white dwarf companion¹⁷. A strong wind is indeed observed^{26,29} in Cyg X-3 and the observed orbital period change agrees with the inferred large mass loss rate in the wind. Even the source X1822-371 has a large and positive rate of change of orbital period². This contrasts with the standard model if the LMXB contains a main sequence companion. As shown in Fig. 1, however, RD-LMXB with a lower main sequence companion may be characterized by a large positive value of $\dot{P}_{\text{orb}}/P_{\text{orb}}$ if $\dot{m}_8 \geq 3$. Alternatively, X1822-371 may contain a white dwarf companion, and Fig. 1 shows that in this case the mass is probably $m \approx (0.1 M_{\odot}) \dot{m}_8$. Both the sign and the magnitude of $\dot{P}_{\text{orb}}/P_{\text{orb}}$ for Cyg X-3 and X1822-371 suggest a stable RD mass transfer with $\frac{1}{3} \leq \beta \leq 1$ and $\Psi \leq 1$. If both Cyg X-3 and X1822-371 contain a very-low-mass white dwarf, it is natural to consider these systems as progenitors of binaries similar to the eclipsing pulsar²⁵ PSR1957+20. In the framework of RD evolution, EXO0748-676 and 4U1820-30 are characterized by a large mass loss from the binary with $\beta \sim \frac{1}{10}$. For both low-mass main sequence and degenerate companions, the RD mass transfer is expected to be unstable with $\Psi \geq 1$, in agreement with the large fluctuations in intensity (by a factor of ~ 5) observed in EXO0748-676 and 4U1820-30. The lower part of Fig. 1 gives the mass range for EXO0748-676 and 4U1820-30 for $\dot{m}_8 \sim 10$ and $\dot{m}_8 \sim 20$, respectively.

The RD model allows two classes of RD-LMXBs, characterized by different values of β . Figure 1 shows that Cyg X-3 and X1822-371 belong to the high- β class, whereas EXO0748-676 and 4U1820-30 belong to the low- β class. New measurements of $\dot{P}_{\text{orb}}/P_{\text{orb}}$ for additional sources would greatly contribute to the determination of the number and classification of RD-LMXBs. At present $\dot{P}_{\text{orb}}/P_{\text{orb}}$ has been measured for four out of fifteen LMXBs with $P_{\text{orb}} \leq 5.5$ h. Of the total sample of ~ 100 LMXBs only ~ 36 binaries have well established orbital periods²⁷. I urge a complete analysis of selection effects affecting the measurements of $\dot{P}_{\text{orb}}/P_{\text{orb}}$, using the four LMXBs in Table 1, to obtain an estimate of the true number of RD-LMXBs. $\square$

Received 3 October 1990; accepted 1 March 1991.

- Kitamoto, S., Miyamoto, S. & Matsui, W. *Publ. astr. Soc. Japan* **39**, 259-285 (1987).
- Hellier, C., Mason, K. O., Smale, A. P. & Kilkenny, D. *Mon. Not. r. astr. Soc.* **244**, 39P-43P (1990).
- Parmar, A. N., Verbunt, F., Smale, A. P. & Corbet, R. H. D. *Astrophys. J.* **366**, 253-260 (1991).
- Tan, J. *et al. Astrophys. J.* in the press.
- Tavani, M. in *Proc. 23rd ESLAB Symp. on X-Ray Binaries* Vol. 1 (ed. White, N.) 241-246 (ESA spec. Publ. 296, 1989).
- Tavani, M. *Bull. Am. astr. Soc.* **24**, 1204 (1989).
- Tavani, M. *Astrophys. J.* **366**, L27-L31 (1981).
- Faulkner, J. *Astrophys. J.* **170**, L99-L102 (1971).
- Rappaport, S., Joss, P. C. & Webbink, R. F. *Astrophys. J.* **254**, 616 (1982).
- Verbunt, F. *Mon. Not. R. Astr. Soc.* **209**, 227-240 (1984).
- Rappaport, S., Verbunt, F. & Joss, P. C. *Astrophys. J.* **275**, 713 (1983).
- Pyllyser, E. & Savonije, G. J. *Astr. Astrophys.* **191**, 57 (1988).
- Stella, L., White, N. & Priedhorsky, W. *Astrophys. J.* **315**, L49 (1987).
- Rappaport, S., Nelson, L. A., Ma, C. P. & Joss, P. C. *Astrophys. J.* **322**, 842-851 (1987).
- Ruderman, M., Shaham, J. & Tavani, M. *Astrophys. J.* **336**, 507 (1989).
- Ruderman, M., Shaham, J., Tavani, M. & Eichler, D. *Astrophys. J.* **343**, 292 (1989).
- Tavani, M., Ruderman, M. & Shaham, J. *Astrophys. J.* **342**, L31 (1989).
- London, R. A., McCray, R. & Auer, L. H. *Astrophys. J.* **243**, 970 (1981).
- Arons, J. *Astrophys. J.* **184**, 539-547 (1973).
- Basko, M. M. & Sunyaev, R. A. *Astrophys. Space Sci.* **23**, 117-128 (1974).
- London, R. A. & Flannery, B. P. *Astrophys. J.* **258**, 260-267 (1982).
- Begelman, M. C. & McKee, C. F. *Astrophys. J.* **271**, 89 (1983).
- Savonije, G. J. in *Accretion-Driven X-Ray Sources* (eds Lewin, W. H. G. & van den Heuvel, E. P. J.) 343 (Cambridge University Press, 1983).
- Molnar, L. A. *Astrophys. J.* **331**, L25 (1988).
- Fruchter, A. S., Stinebring, D. R. & Taylor, J. H. *Nature* **333**, 237 (1988).
- Molnar, L. A., Kouba, J. A. & Raymond, J. C. *Bull. Am. astr. Soc.* **22**, 1339 (1990).
- Ritter, H. *Astr. Astrophys. Suppl.* (in the press).
- van der Klis, M. & Bonnet-Bidaud, J. M. *Astr. Astrophys.* **95**, L5-L9 (1981).
- Willingdale, R., King, A. R. & Pounds, K. A. *Mon. Not. r. astr. Soc.* **215**, 295 (1985).

ACKNOWLEDGEMENTS. I thank Professors L. Stella, M. van der Klis and J. van Paradijs for valuable discussions and the communication of results concerning 4U1820-30. This work was performed at IGGP-LLNL under a US Department of Energy contract.

The Sun's luminosity over a complete solar cycle

Richard C. Willson* & Hugh S. Hudson†

* Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91109, USA

† University of California at San Diego, La Jolla, California 92093, USA

THE Active Cavity Radiometer Irradiance Monitor (ACRIM I), an instrument carried on NASA's Solar Maximum Mission satellite, measured the Sun's luminosity (total power outflow) from early 1980 to late 1989¹⁻⁵. Here we present the first account of the complete ACRIM I data set, and give evidence confirming our previous suggestion that solar luminosity varies with the 11-year solar cycle⁶. As previously reported, this slow variation closely follows statistical measures of the distribution of magnetic and photospheric features on the Sun's surface⁴⁻⁸. But there was an exception to this correlation in the form of a remarkable irradiance excess during 1980, at about the time of the sunspot maximum of solar cycle 21. The linkage, over a whole cycle, of luminosity variation to photospheric activity suggests the existence of an unknown physical mechanism other than the thermal diffusion model that explains luminosity deficits due to sunspots. Luminosity models connecting total irradiance to global indicators of solar activity, such as the equivalent width of the 1,083-nm helium line, are consistent with the gross features of the variability, but fail to account for the 1980 irradiance excess.

Nearly a decade of high-precision monitoring of solar luminosity by the ACRIM I experiment ended with re-entry of the Solar Maximum Mission (SMM) spacecraft in December 1989. Significant solar variability was found on all timescales, ranging from individual samples (1.024 seconds) to the full duration of the observations (9.75 years). These results were the first to demonstrate luminosity fluctuations associated with a wide variety of solar phenomena: sunspots, active regions, oscillations, granulation and the bright network¹⁻¹⁰, a non-uniform, filamentary distribution of excess emission outside active regions.

Here we discuss the most important aspects of the variation in solar luminosity on timescales of a year or more observed by ACRIM I during solar cycles 21 and 22. Such variations must be linked to mechanisms deep within the convection zone or even below, where the source of the solar magnetic cycle is thought to reside. The successes and failures of several empirical models of irradiance will be discussed below^{7,8,11}.

The results of the ACRIM I experiment, normalized to 1 AU, are shown as daily averages in Fig. 1. The number of independent

samples was usually near the maximum available of 10⁴ per day during the two periods of precision solar pointing. Daily mean values had an average standard error of 23 parts per million (p.p.m.) of the total flux during the first such period (February to November 1980) and 14 p.p.m. during the second period (after March 1984), reflecting the intrinsic solar variability which was higher during the first of these two experimentally equivalent periods. During the intervening SMM spin-stabilized mode of observation, only ~100 independent samples were acquired per day, resulting in a higher average standard error of the daily means of 76 p.p.m.

The irradiance declined systematically by ~0.1% from launch through mid-1986 to a minimum in near-coincidence with the minimum in solar activity in September 1986. A rapid increase followed, corresponding to the build-up of solar activity in cycle 22.

The general pattern of these results is a variation of the total irradiance in step with the solar magnetic activity cycle, but with considerable short-term fluctuation superimposed on it. The rapid fluctuations on timescales of a few days have been identified with sunspots, and those of a few months, with faculae¹⁻⁵. The data set as a whole strongly suggests a true variation of solar luminosity with the 11-year solar cycle.

It has been suggested previously^{7,8,11} that the initial rapid decrease of the total solar irradiance was due to degradation of the sensor during this period. Regression models based on indicators of solar magnetic activity (He 1,083 nm, H Lyman- α and Fraunhofer lines) predict significantly lower values of irradiance than those measured by ACRIM I and an independent experiment, the Nimbus-7/ERB (Earth Radiation Budget)^{12,13}. We establish here that (1) the internal calibration of ACRIM I is sound during this interval the good agreement with Nimbus-7/ERB results confirms this calibration; and the mismatch with the regression-based models in 1980 probably indicates the existence of a new feature of the solar cycle that cannot be seen in the previously known indicators.

The 81-day running means of the irradiance measured by ACRIM I and that predicted by a regression model based on the He 1,083-nm flux are shown in Fig. 2. The measured irradiance is corrected for sunspots by the 'photometric sunspot index' (PSI). The model is given as equation (1), where S_{model} is the predicted irradiance (in W m^{-2}) and EW_{He} is the equivalent width of the solar He line at 1,083 nm.

$$S_{\text{model}} = 1,364.08 + 0.0636161 \times EW_{\text{He}} \quad (1)$$

The plot, in units of percentage variation relative to the means of the time series, shows general agreement between the measured and modelled irradiances except for 1980. The divergence there is significant: the average values for 1980 are 0.093%

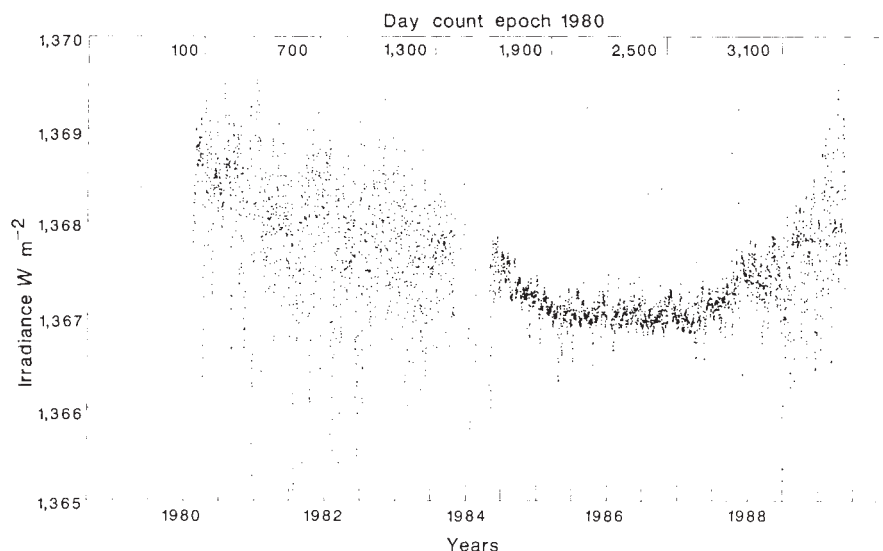


FIG. 1 Total solar irradiance observations by the ACRIM I experiment on the Solar Maximum Mission (SMM). Points are the mean values for a day's observations.

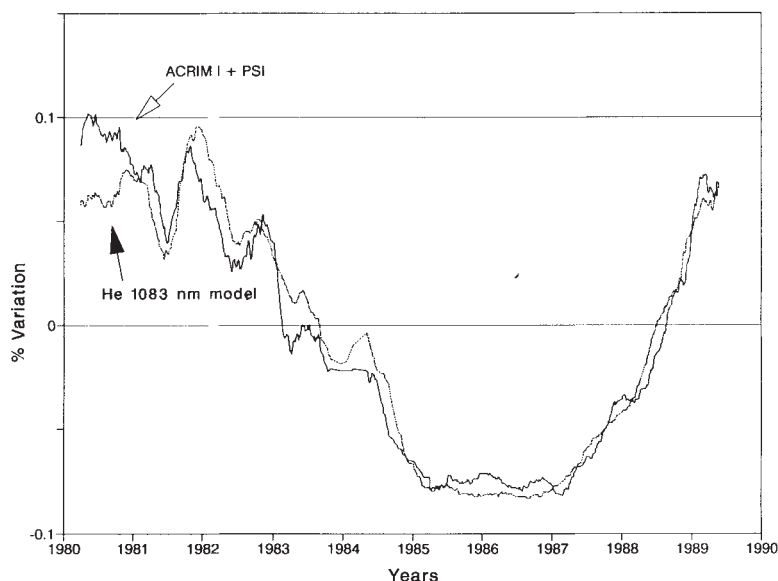


FIG. 2 Percentage variation of 81-day running means for the ACRIM I observations, corrected for sunspot effects by the photometric sunspot index (solid line), and the He 1,083-nm equivalent width regression model of irradiance (dashed line).

for ACRIM I+PSI and 0.063% for the He-1,083 model, an average difference of 0.03%.

Two sources of experimental evidence help us to evaluate the significance of the divergence. The first is another total solar irradiance experiment, the Nimbus 7/ERB mentioned above. We compared the ACRIM I and ERB data after correcting for sunspot effects by using the PSI to remove the irradiance 'sunspot deficit' sensitivity not present in the He 1,083-nm flux³. (The ACRIM I and ERB data are sensitive to 'sunspot deficit', whereas the He 1,083-nm flux is not.) The average ACRIM I-ERB residual in 1980 was 0.002%. The irradiance excess of average amplitude 300 p.p.m. in 1980 is therefore 'resolved' by the agreement of ACRIM I and ERB in 1980 to within 20 p.p.m. There is therefore strong experimental evidence for the irradiance excess.

The second source of evidence is an internal calibration system within ACRIM I itself which was sufficient to rule out significant uncorrected degradation at the level of variability detected so far¹. This calibration is based on comparisons of the three nominally identical active cavity radiometer sensors, which are independently shuttered. The first (A) was used continuously as the solar monitor. Shutters on the second two (B and C) were operated infrequently to calibrate the degradation of the first caused by exposure to the Sun (Fig. 3). Calibration comparisons

were made between sensors A and B at roughly monthly intervals, less frequently between all three sensors. The primary sensor's total degradation of 600 p.p.m. over the 9.75-year mission was calibrated with a residual uncertainty of <50 p.p.m. using the ratio of sensors A and C. Eventual degradation of sensor B, the more frequently used calibration reference, can be seen after day 2,500.

A procedure designed to detect and calibrate the initial degradation of ACRIM I's primary sensor¹ found that 85% of the total degradation in 1980 (~180 p.p.m.) occurred in the first 100 days of observations, from mission days 62 to 163 (Fig. 3). The residual uncertainty of this calibration is better than ± 12 p.p.m., the same order of magnitude as the agreement between ACRIM I and ERB. The 300-p.p.m. 1980 irradiance excess is therefore defined by two independent experimental methods to have an uncertainty of ≈ 20 p.p.m. We believe the variations observed by ACRIM I, including the disputed 1980 irradiance excess, are unambiguously solar in origin.

The ACRIM I data show two principal features of solar luminosity variation on long timescales: the magnetic cycle itself and the irradiance excess of 1980. From the strict standpoint of time-series analysis, the apparent luminosity variation during the solar cycle must await confirmation by observations during future solar cycles. But solar proxy models may be useful here:

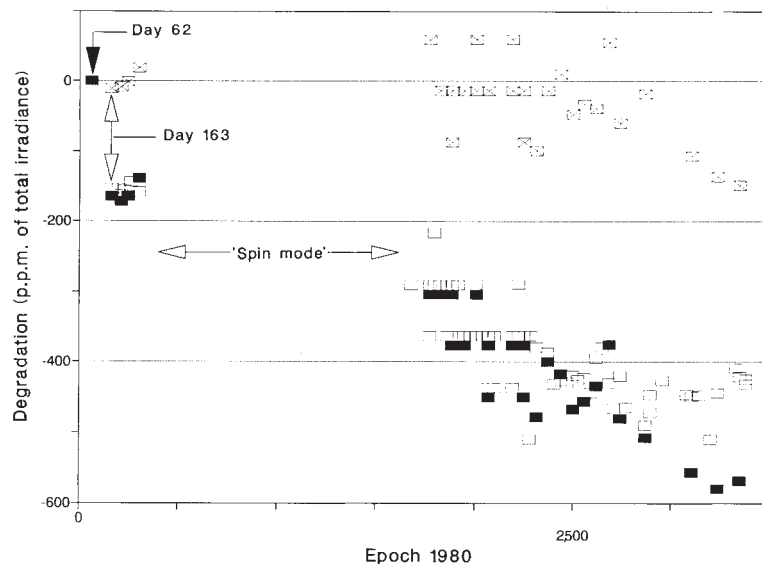


FIG. 3 Results of the ACRIM I degradation calibration experiment plotted as the ratios of its sensors A, B and C. Empty squares are the ratio A/B, solid squares the ratio A/C and crossed squares the ratio B/C. The gap between days 346 and 1,585 results from insufficient sample acquisition during SMM's 'spin mode'.

the gross features of solar cycle behaviour predicted by the He 1,083-nm model matched those of the measured irradiance reasonably well (except for 1980). The data presented here, in this context, indicate that future solar cycles will have similar features which can be detected by sufficiently precise total irradiance observations.

The 1980 irradiance excess is more problematic, as it does not correspond with other indicators of solar activity. The evolution of solar luminosity over the next decade, as chronicled by the ERB, the upcoming and planned satellite missions will be exciting to watch. We should soon have results from the Nimbus-7/ERB experiment with which we can test the prediction made here: of a repetition of the 1980 irradiance excess in 1990–91. □

Received 13 August 1990; accepted 11 March 1991.

- Willson, R. C., Gulkis, S., Janssen, M., Hudson, H. S. & Chapman, G. A. *Science*, **211**, 700–702 (1981).
- Willson, R. C. *J. geophys. Res.* **86**, 4319–4326 (1982).
- Hudson, H. S., Silva, S., Woodard, M. & Willson, R. C. *Solar Physics*, **76**, 211 (1982).
- Willson, R. C., Hudson, H. S., Frohlich, C. & Brusa, R. W. *Science* **234**, 1114–1117 (1986).
- Foukal, P. & Lean, J. *Astrophys. J.* **302**, 826–835 (1986).
- Willson, R. C. & Hudson, H. S. *Nature* **332**, 810–812 (1988).
- Foukal, P. A. & Lean, J. L. *Astrophys. J.* **328**, 347 (1988).
- Foukal, P. & Lean, J. *Science* **246**, 556–558 (1990).
- Woodard, M. & Hudson, H. *Nature* **305**, 589 (1983).
- Woodard, M. & Noyes, R. *Nature* **318**, 449 (1985).
- Livingston, W. C., Wallace, L. & White, O. R. *Science* **240**, 1765–1767 (1988).
- Hickey, J. R., Alton, B. M., Kyle, H. L. & Hoyt, D. V. *Space Sci. Rev.* **48**, 321–342 (1989).
- Hoyt, D. V. & Kyle, H. L. *An Alternative Derivation of the Nimbus 7 Total Solar Irradiance Variations in Climate Impact of Solar Variability NASA Conf. Publ. 3086* (eds Schatten, K. & Arking, A.) 293–300 (NASA Goddard Space Flight Center, Greenbelt, 1990).

ACKNOWLEDGEMENTS. We thank J. Harvey (Kitt Peak Observatory) for the He 1,083-nm data and D. Hoyt (NASA/GSFC) for the revised Nimbus-7/ERB results. The SMM/ACRIM I experiment is supported by NASA at JPL, Cal. Tech. Work at UCSD was supported by NASA.

A collection of diverse micrometeorites recovered from 100 tonnes of Antarctic blue ice

M. Maurette*, C. Olinger†**, M. Christophe Michel-Levy‡, G. Kurat§, M. Pourchet||, F. Brandstätter§ & M. Bourot-Denise¶

* CSNSM, IN2P3-CNRS, Bâtiment 108, 91405 Campus-Orsay, France

† McDonnell Center for the Space Sciences and Physics Department, Washington University, St Louis, Missouri 63130, USA

‡ Laboratoire de Minéralogie-Cristallographie, Université Paris-VI, 75252, Paris, France

§ Naturhistorisches Museum, Box 417, A-1014 Vienna, Austria

|| Laboratoire de Glaciologie, B.P. 96, 38402 St-Martin d'Hères, France

¶ Laboratoire de Minéralogie, Muséum d'Histoire Naturelle, 75005 Paris, France

STUDIES of meteorites and interplanetary dust particles (IDPs) have provided constraints on the formation and evolution of the Solar System¹, and have identified pre-solar interstellar grains^{2–4}. Here we describe a new type of meteoritic material, intermediate in size between meteorites and IDPs. Melting and filtering of ~100 tonnes of blue ice near Cap Prudhomme, Antarctica, yielded ≥7,500 irregular, friable particles and ~1,500 melted spherules, ~100 µm in size, both showing a 'chondritic' composition suggestive of an extraterrestrial origin^{5,6}. For the present work, we analysed the composition and texture of 51 irregular particles and 25 spherules. The irregular particles appear to be unmelted, and have similarities with the fine-grained matrix of primitive carbonaceous chondrites, but are extremely diverse in composition. Isotopic analysis of trapped neon confirms an extraterrestrial

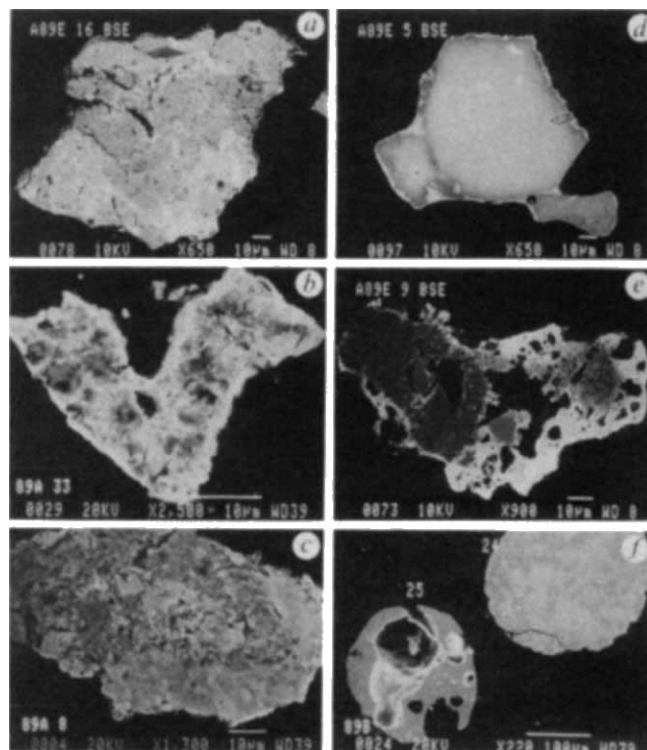


FIG. 1 Scanning electron micrographs of polished sections of ~100 µm sized Antarctic gas rich micrometeorites and spherules. *a*, Very fine-grained unmelted micrometeorite. *b*, Unmelted particle containing observable phyllosilicates and crystals that are too small for individual electron microprobe analysis. *c*, Clast-bearing micrometeorite. *d*, Coarse-grained particle composed of two large olivine crystals, with a small crystal of pyroxene protruding on the lower right corner. *e*, Coarse-grained, partially melted scoria particle with large relict olivines. *f*, Typical vesicular glassy spherule with no relict crystals (left-hand side).

origin for 16 of 47 irregular particles and 2 of 19 spherules studied, and strongly suggests that they were exposed in space as micrometeoroids. These large Antarctic micrometeorites constitute a new family—or at least a new population—of Solar System objects, in a mass range corresponding to the bulk of extraterrestrial material accreted by the Earth today.

In January 1988, two steam generators delivering pressurized jets of water at 80 °C were used to melt ~40 pockets of ice with volumes of 2–3 m³ each, ~6 km from the French station of Dumont d'Urville, near Cap Prudhomme, in Antarctica. About 5–10 tonnes of water were pumped and filtered daily on a stack of three stainless steel sieves with openings of 50, 100 and 400 µm. The whole operation yielded ~10 g of dust of diameter ≥50 µm including a total number of ~1,500, ~5,000, and 9 spherules in the 50–100-µm, 100–400-µm and ≥400-µm fraction, respectively. Closer scrutiny in the laboratory gave the ratio of suspected micrometeorites to melted spherules as ≥5 and ~0.3 in the two smaller fractions, respectively.

Seventy apparently unmelted grains were hand-picked from the 50–100-µm fraction, relying on simple criteria such as the particle having a dark grey-to-black colour and irregular shape. Additionally, 30 spherules were selected for study in the same size range. Polished sections were analysed with a scanning electron microscope, equipped with an energy dispersive X-ray spectrometer. Fifty irregular 'stony' grains and all but one of the stony spherules could be classified as 'probably extraterrestrial' relying on the observation that their silicon, magnesium, iron and aluminium contents occurred in roughly chondritic proportions^{5,6}. These proportions are similar to those previously obtained for deep sea and Greenland spherules^{7,8} containing a high concentration of the cosmogenic nuclides ¹⁰Be and ²⁶Al

** Present address: Los Alamos National Laboratory, INC-7 MS J514, Los Alamos, New Mexico 87545, USA.

and relatively unmelted Greenland micrometeorites⁶ containing implanted solar neon. One additional particle has an apparently extraterrestrial metallic Fe–Ni composition. Following optical and scanning electron microscopy, electron microprobe analyses were performed on bulk samples as well as on individual phases, clasts and matrices to determine major and minor element contents (for oxide compositions, see Supplementary Information).

Figure 1 illustrates some textural features of the 50 stony particles. The most common are unmelted particles, composed of fine-grained phyllosilicates with variable amounts of poorly characterized (Fe–Ni) oxide phases, olivines, low-calcium pyroxenes, high-calcium pyroxenes (diopside and fassaite), and rare grains of Fe–Ni metal (Fig. 1a to c). A few unmelted particles consist mostly of olivine (Fig. 1d). There is a range from unmelted particles to partially melted scoria particles, which consist mainly of glass instead of phyllosilicates (Fig. 1e). About one-sixth of the 'probably extraterrestrial' particles in the 50–100- μm range are melted spherules (Fig. 1f). All selected irregular particles were at least partially covered by a thin rim of magnetite.

Figure 2 compares the bulk major element contents of unmelted and partially melted scoria particles (solid line histogram) with the average composition of primitive carbonaceous chondrites (C1)⁹. From the peaks in elemental ratio distributions, the compositions of micrometeorites look 'approximately chondritic' in Mg, Al, Si and Fe, but with a wide dispersion of elemental ratios. Depletions are observed, however, in Ca, Ni and S relative to chondritic values. Shaded areas represent analyses of the matrices of 27 carbonaceous chondrites (3 C1, 11 C2 and 13 C3), and of matrix and chondrule rims from one unequilibrated ordinary chondrite (Tieschitz). The dispersion observed in element/Si ratios for Mg, Al and Fe in the matrices of carbonaceous and unequilibrated ordinary chondrites is of the same order as that observed for the unmelted particles. Similarly, the strong depletion of calcium is of the same order when the matrix is considered instead of the bulk. This depletion of Ca in the matrices of carbonaceous chondrites was recognized earlier by McSween and Richardson¹⁰, who also found a depletion in S but no depletion in Ni.

The chemical composition of non-porous areas of fine-grained micrometeorites, as well as infrared absorption bands (1,580 and 3,260 cm^{-1}) of 12 similar unmelted particles¹¹, supports the existence of hydrous silicates and oxides in most of these

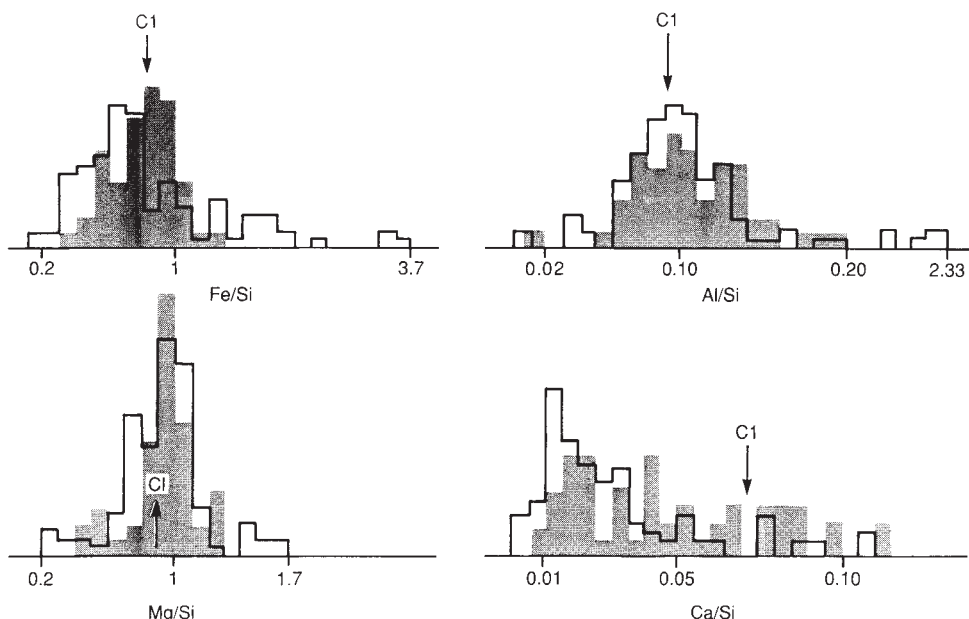
particles. However, most Antarctic micrometeorites also contain anhydrous magnesium-rich olivines and pyroxenes. The coexistence of hydrous and anhydrous minerals indicates a similarity between micrometeorites and C2 carbonaceous chondrites. It also suggests some relationships with much smaller stratospheric IDPs. This view is supported by the frequency distributions of elemental ratios (Fig. 2), which are rather similar to those measured for hydrous IDPs¹².

Phase compositions, determined by electron microprobe analysis, independently support the extraterrestrial origin of the unmelted and partially melted scoria particles, as well as their relationship with carbonaceous chondrites. Phyllosilicate compositions are similar to those observed in C2 chondrites¹⁰, with roughly chondritic abundances of Al, Ti, Cr and Mn, sub-chondritic abundances of Ni, Ca, Na and K, and highly variable FeO/MgO ratios (0.28–3.3 by weight). Olivines also have variable FeO/MgO ratios ($F_{\text{O}_{99}}-F_{\text{O}_{25}}$) with magnesian compositions predominating. Their minor element contents are usually high, with ranges of 0.1–1.0 wt% and 0.02–2.8 wt% for Cr_2O_3 and Al_2O_3 , respectively, and most olivines have Fe/Mn ratios < C1 values (olivines enriched in Mn). These features are typical for olivines from carbonaceous chondrites^{13–15} and IDPs¹⁶ and are not compatible with either equilibrated ordinary chondrites or terrestrial olivines.

Pyroxenes are present in a variety of compositions. Low-calcium pyroxenes are mostly enstatites (iron-poor) and are rich in minor elements (Al, Cr and Mn). Of the calcium-rich pyroxenes, two varieties are present: diopside (rich in Al, Ti and Mn) and fassaite, a refractory pyroxene very rich in Al_2O_3 (16 wt%) and TiO_2 (0.7 wt%). All pyroxene compositions are typical for those found in carbonaceous chondrites and IDP pyroxenes^{15,17}. The metal is nickel-rich (35 wt% Ni) and has a primitive Ni/Co ratio (~ 22) clearly indicating a primitive heritage (ordinary chondrites have fractionated metal compositions). Poorly characterized (probably hydrous) oxides are commonly nickel-bearing and have Ni/Co ratios similar to bulk C1 chondrites (~ 23). Low-nickel pentlandite (3.8–14.7 wt% Ni) is the main sulphide phase, also typical for carbonaceous chondrites and IDPs¹⁷. Some particles, like 89A-33 (Fig. 1b), which do not seem to have been melted, contain a glassy phase with a roughly chondritic composition which is depleted in nickel (and probably other siderophile elements). Such glasses are rare in C2 meteorites¹⁸, but common in IDPs¹⁹.

Noble gas analysis was performed using low-blank laser

FIG. 2 Ratios of major element to silicon for Cap Prudhomme unmelted micrometeorites (full-line histograms) compared with matrices of carbonaceous chondrites and of the unequilibrated ordinary chondrite Tieschitz (shaded histograms) according to McSween and Richardson¹⁰ and our own unpublished data. Multiple analyses were performed on micrometeorites and the carbonaceous chondrites at scales of $\sim 10\ \mu\text{m}$ (our data) and $100\ \mu\text{m}$ (ref. 10) to assess heterogeneity and facilitate comparison with IDPs.



extraction techniques²⁰ on 19 unmelted and 28 partially melted chondritic particles considered 'probably extraterrestrial' as well as on 9 non-chondritic particles and 19 chondritic spherules. A total of 26 particles contained enough neon to permit isotopic measurements (for neon data, see Supplementary Information). Eighteen of these samples (9 unmelted, 7 partially melted scoria and 2 spherules) have isotopic compositions at least 2σ different from terrestrial neon, confirming their extraterrestrial origin (Fig. 3). (The others may also be extraterrestrial, but cannot be unequivocally confirmed as such.) Most of the gas-rich particles have neon isotopic compositions distinct from air, containing a mixture of neon from solar wind²¹ and from solar energetic particles²². In addition, several particles have compositions displaced towards higher $^{21}\text{Ne}/^{22}\text{Ne}$ values, as would be expected for spallation neon, which is produced by cosmic-ray interactions with the silicates. The low, but measurable, spallation excesses observed in some grains can be attributed to irradiation by cosmic rays either when the micrometeorites existed as small particles in space or, possibly, during a previous residence time in the near-surface region of a larger parent body.

Particles analogous to Antarctic fine-grained unmelted particles are not generally observed in collections of less friable

and more heavily weathered micrometeorites collected from Greenland²³ and deep-sea sediments²⁴. In spite of these differences, isotopic results for all of these particles are nearly identical^{6,25,26} and are similar to neon compositions measured in stratospheric IDPs²⁷. This may reflect a similarity in their exposures to energetic particles in space.

Many of these particles have surprisingly high concentrations of implanted solar neon. Over half of the samples confirmed to be extraterrestrial by neon analysis have neon concentrations in excess of $10^{-5} \text{ cm}^3 \text{ g}^{-1}$ at STP. This contrasts sharply with larger meteorites, where all but five 'gas-rich' meteorites have neon concentrations below this value, and two of those are lunar meteorites²⁸. These high concentrations, as well as the observation that most of the neon is implanted solar gas, strongly suggest that most of the samples were exposed in space as micrometeoroids (they are not simply fragments of larger meteorites produced by ablation on entering the atmosphere).

Neon analysis confirms an extraterrestrial origin for only $\sim 1/4$ of the 'chondritic' particles studied, but these include representative samples of unmelted particles, partially melted scoria particle and, surprisingly, two of the melted cosmic spherules. Moreover, based on mineralogical, chemical and textural grounds (Fig. 1), gas-rich and gas-poor particles identified as 'probably extraterrestrial' are indistinguishable, except for a higher content of anhydrous mineral inclusions $\geq 5 \mu\text{m}$ in the gas-rich particles. It is therefore likely that most or all of the gas-poor chondritic particles are also micrometeorites.

We have given elsewhere²⁹ a rough estimate of the flux of micrometeorites (size range $50\text{--}300 \mu\text{m}$) as recovered from either the Greenland or the Antarctica ice sheets, which agrees within a factor ≈ 3 with the total mass influx on the Earth ($\sim 20,000 \text{ tonne yr}^{-1}$). The Antarctic ice provides an inexhaustible source of micrometeorites that show an amazing diversity (no two grains are exactly alike in this set). Overall, in chemistry and mineral composition they have strong similarities with carbonaceous chondrites and IDPs. They are not, however, identical. Unmelted and partially melted scoria particles are depleted in both nickel and sulphur relative to carbonaceous chondrites and to IDPs¹². One unmelted particle (89A-36) contains an olivine crystal with a NiO content ($\sim 0.4\%$) unknown in meteoritic olivines. If further work shows that such differences do not reflect weathering in the ice or frictional heating in the atmosphere, then they may represent a new population of Solar System objects. Further studies of these unique particles could broaden our knowledge of Solar System objects accreting to Earth today and may well provide new insights into nebular and pre-nebular processes. □

Received 13 December 1990; accepted 8 April 1991.

- Kerridge, J. F. & Matthews, M. S. (eds) *Meteorites and the Early Solar System* (University of Arizona Press, Tucson, 1988).
- Lewis, R. S., Ming, T., Wacker, J. F., Anders, E. & Steel, E. *Nature* **326**, 160–162 (1987).
- Bernatowicz, T. et al. *Nature* **330**, 728–730 (1987).
- Amari, S., Anders, E., Virag, A. & Zinner, E. *Nature* **345**, 238–240 (1990).
- Maurette, M., Jéhanno, C., Robin, E. & Hammer, C. *Nature* **328**, 699–702 (1987).
- Olinger, C. T., Maurette, M., Walker, R. M. & Hohenberg, C. M. *Earth planet. Sci. Lett.* **100**, 77–93 (1990).
- Raisbeck, G. M. & Yiou, F. *Meteoritics* **22**, 485–486 (1987).
- Raisbeck, G. M. & Yiou, F. *Meteoritics* **24**, 318 (1989).
- Anders, E. & Ebihara, M. *Geochim. cosmochim. Acta* **46**, 2363–2380 (1982).
- McSween, H. Y. & Richardson, S. M. *Geochim. cosmochim. Acta* **41**, 1145–1161 (1977).
- Maurette, M., Beccard, B., Bonny, P. & Christophe Michel-Levy, M., *Proc. 24th ESLAB Symp.* (ed. Batrack, B.) (Eur. Space Ag., in the press).
- Schramm, L. S., Brownlee, D. E. & Wheelock, M. M. *Meteoritics* **24**, 99–112 (1989).
- Steele, I. M. *Geochim. cosmochim. Acta* **50**, 1379–1395 (1986).
- Kurat, G., Mayr, M., Ntafos, T. H. & Graham, A. L. *Meteoritics* **24**, 35–42 (1989).
- Hoinkes, G. & Kurat, G. *Meteoritics* **10**, 429–430 (1975).
- Klöck, W., Thomas, K. L., McKay, D. S. & Palme, H. *Nature* **339**, 126–128 (1989).
- Tomeoka, K. & Buseck, P. R. *Earth planet. Sci. Lett.* **69**, 243–254 (1984).
- Hoinkes, G. & Kurat, G. *Meteoritics* **14**, 423 (1979).
- Thomas, K. L., Zolensky, M. E., Klöck, W. & McKay, D. S. *Lunar planet. Sci.* **21**, 1250–1251 (1990).
- Hohenberg, C. M., Nichols, R. H. Jr, Olinger, C. T. & Goswami, J. N. *Geochim. cosmochim. Acta* **54**, 2133–2140 (1990).
- Geiss, J., Buehler, F., Cerutti, H., Eberhardt, P. & Filleux, C. *Apollo 16: Preliminary Science Report* NASA SP-315, 14.1–14.10 (1972).
- Wieler, R., Baur, H. & Signer, P. *Geochim. cosmochim. Acta* **50**, 1997–2017 (1986).
- Robin, E., Christophe Michel-Levy, M., Bourot-Denise, M. & Jéhanno, C. *Earth planet. Sci. Lett.* **97**, 162–176 (1990).

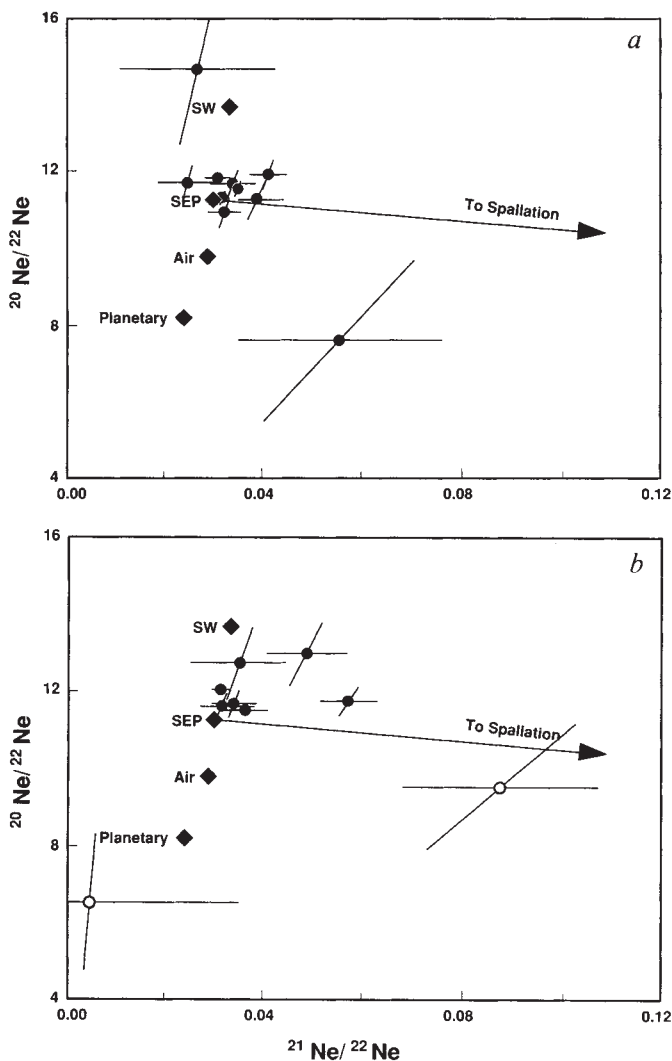


FIG. 3 Isotopic composition of neon extracted from Antarctic micrometeorites, with 1σ uncertainties. a, Unmelted particle (●). b, Partially melted scoria particles and melted spherules (open circles). Diamonds indicate the compositions of solar wind (SW), solar energetic particle (SEP), atmospheric (Air) and planetary neon. Most Antarctic particles with measurable neon are dominated by the surface-implanted solar components, demonstrating their extraterrestrial origin and suggesting that they are exposed in space as micrometeoroids.

24. Brownlee, D. E. *Ann. Rev. Earth planet. Sci.* **13**, 147–173 (1985).
 25. Amari, S. & Ozima, M. *Geochim. cosmochim. Acta* **52**, 1087–1095 (1988).
 26. Fukumoto, H., Nagao, K. & Matsuda, J. I. *Geochim. cosmochim. Acta* **50**, 2245–2253 (1986).
 27. Nier, A. O. & Schlutter, D. J. *Meteoritics* **25**, 263–267 (1990).
 28. Schultz, L. & Kruse, H. *Meteoritics* **24**, 155–172 (1989).
 29. Maurette, M., Hammer, D. & Pourchet, M. in *From Mantle to Meteorites* (eds Gopalan, K., Gaur, V. K., Somayajulu, B. L. & MacDougall, J. D.) (Indian Academy of Sciences, Bangalore, 1990).

SUPPLEMENTARY INFORMATION. Requests should be addressed to the London editorial office of *Nature*.

ACKNOWLEDGEMENTS. We thank R. M. Walker, D. Brownlee and C. M. Hohenberg for helpful discussions, and P. Galle and P. Siry for help. Funding for the Cap Prudhomme expedition was granted by 'Expéditions Polaires Françaises' and 'Terres Australes et Antarctiques Françaises'. Financial support is acknowledged from IN2P3 and INSU in France, 'Fonds zur Förderung der wissenschaftlichen Forschung' in Austria, and NASA.

Structural origin of reduced critical currents at $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ grain boundaries

M. F. Chisholm & S. J. Pennycook

Solid State Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831-6024, USA

THE critical current density across individual grain boundaries in thin films of the high- T_c superconductor $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ (YBCO) has been found^{1–4} to be inversely proportional to lattice misorientation for tilts up to $\sim 10^\circ$. Reports of impurity segregation^{5,6} at grain boundaries, and variations in the chemical stoichiometry^{7,8}, have led to the view that deviations from the ideal composition are responsible for the depressed superconducting order parameter at the boundary. Here we present images of YBCO grain boundaries obtained by a scanning transmission electron microscope in Z -contrast mode^{9,10}, which show that chemical segregation does not necessarily occur at these boundaries. A simple model of the strain associated with the grain-boundary dislocations provides a reasonable physical explanation of the suppressed superconductivity. The surprisingly large effect of strain implied by our model has implications beyond critical currents, for the physics and applications of any thin-film YBCO structures involving strained epitaxial layers.

In epitaxial thin films of $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ on polycrystalline SrTiO_3 , the critical current density across a grain boundary decreases drastically with increasing misorientation angle,^{1–4} and the boundary behaves as a weak link for misorientations greater than 10° . Recent results from bulk-scale bicrystals, however, have indicated that certain high-angle grain boundaries do not exhibit this weak-link response¹¹. No satisfactory microscopic explanation for the origin of this behaviour of grain boundaries has so far been proposed^{3,12}. Additionally, it has structural models of low-angle boundaries do not lead to the observed angular dependence of the transport critical current^{2,4,12}. The most commonly accepted view is that deviations from the ideal stoichiometry are responsible for the low critical current densities¹³. There are several reports of compositional variations at grain boundaries^{5–8}, but only small decreases in already low critical current densities have been directly correlated with deviations from the stoichiometric composition in the boundary regions⁸.

We have used the Z -contrast technique^{9,10} for forming chemically sensitive high-resolution images in a scanning transmission electron microscope (VG Microscopes HB501UX STEM operating at 100 kV) to investigate the composition of YBCO grain boundaries. In this technique, a fine electron beam of width 0.22 nm (full width at half maximum intensity) is scanned across the sample, and the electrons scattered at high angles are collected by an annular detector and used to form an image. The high-angle scattering is proportional to Z^2 , where Z is the atomic number of the species under the probe. Thus the image can be thought of as a map showing at atomic resolution the scattering

power of the sample^{9,10}. This technique can provide analysis of regions of the sample down to a size comparable to the superconducting coherence length. We have examined boundaries of the same type as those investigated in refs 1–4. Such tilt boundaries consist of an array of uniformly spaced edge dislocations produced to accommodate the mismatch¹⁴. Our grain-boundary samples were taken from polycrystalline thin films of $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ grown on Y_2O_3 -stabilized ZrO_2 and from a pellet of sintered powder. Figure 1 shows Z -contrast images of a 13° tilt boundary at which the adjacent grains nearly share a common $[100]$ direction. In Fig. 1a the beam passes parallel to the crystal planes of both grains, which are seen to be separated by an array of triangular defects. The planes of barium atoms ($Z = 56$) show higher intensity than the yttrium ($Z = 39$) planes, and terminate both grains along the long sides of the triangular zones. Phase-contrast imaging using the STEM bright-field detector clearly shows that these faceted dislocation cores are amorphous. To determine whether the amorphous material has the same composition as the crystalline superconductor, it is necessary to tilt the right-hand grain by a few degrees so that the clear image of the crystal planes is lost. The intensity scattered by the crystal in a random orientation will equal the intensity scattered by a random (amorphous) arrangement having the same overall composition. The contrast in the annular detector signal can be quantified very simply using appropriate screened cross-sections for elastic scattering¹⁵. Thus we can probe any segregation to the amorphous zones with a sensitivity limited by the image statistics and by the extent to which we can avoid thickness variations and residual channelling effects, which we estimate to be $\pm 2\%$.

Within this limit no contrast changes are detectable between the amorphous zones and the randomly oriented crystal (Fig. 1b). Similar results were obtained from other tilt boundaries, including a 34° high-angle (high-energy) tilt boundary. This rules out the presence of any of the likely phases, including $\text{Y}_2\text{Ba}_4\text{Cu}_7\text{O}_{14}$, $\text{YBa}_2\text{Cu}_4\text{O}_8$ and BaCuO_2 , and sets limits to the range of stoichiometry that would not be detectable as

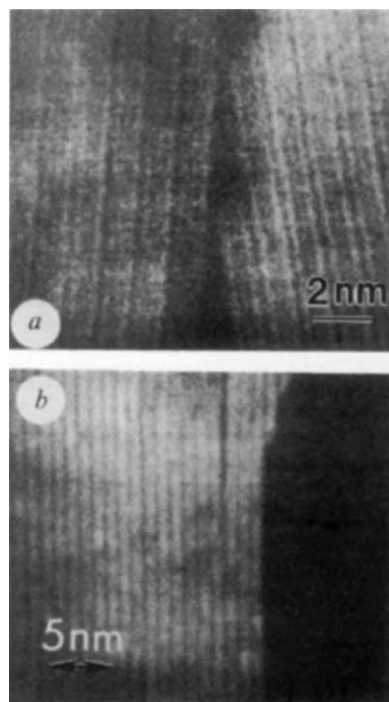


FIG. 1 Z -contrast images of a 13° $[100]$ tilt boundary. *a*, Both grains oriented for electron channelling along the (001) planes, showing an array of triangular amorphous zones at the boundary. *b*, Sample tilted so that the right-hand grain is not channelling, showing no detectable contrast variations along the grain boundary.

$\text{Y}_{0.87 \leq x \leq 1.13} \text{Ba}_{1.92 \leq x \leq 2.07} \text{Cu}_{2.79 \leq x \leq 3.21} \text{O}_{4.91 \leq x \leq 9.09}$, where only one component is assumed to vary at a time. The method is rather insensitive to oxygen because of its relatively low atomic number, but given the high mobility of oxygen at low temperatures¹⁶ it seems doubtful that any microscopic measurement on a thin specimen could provide meaningful oxygen contents. We estimate the spatial resolution of our measurement to be 1 nm, significantly better than has been achieved using X-ray fluorescence. This resolution is possible because the number of elastically scattered electrons detected is typically 10^5 times greater than the number of X-rays collected from the same area, allowing us to work with thicknesses of only 10–20 nm, which greatly reduces beam-broadening effects. Our results therefore represent the most sensitive measurements so far of grain-boundary composition. They indicate clearly that, although compositional variations undoubtedly can occur, they are not necessarily present.

In the light of these results, we therefore re-examine structural models of grain boundaries and attempt to understand how a dislocation array can cause a large reduction in critical current. We start with the simplest model and assume that the order parameter for superconductivity is depressed only within a strained region associated with the grain-boundary dislocations, which we assume has a well defined radius, r_m , in the plane of the boundary. Therefore, the ratio of the critical current density across a grain boundary, J_c^{gb} , to that within the grain, J_c^G , is $(D - 2r_m)/D$, where $D = |\mathbf{b}|/\sin \theta$ is the dislocation spacing, $\mathbf{b}$ is the Burgers vector of the dislocation, and θ is the tilt angle of the two adjacent grains. For small tilt angles,

$$J_c^{gb}/J_c^G = 1 - \frac{2r_m}{|\mathbf{b}|} \theta \quad (1)$$

That is, the normalized grain-boundary critical current density decreases linearly with θ for small misorientations. In Fig. 2 we show a linear plot of the data of Dimos *et al.*⁴, which clearly shows this predicted linear behaviour in the small-angle regime. From a least-squares fit of the slope between $\theta = 0^\circ$ and 10° , we determine $r_m \approx 2.9|\mathbf{b}|$, which is much larger than the value assumed previously^{2,6}.

We believe that the physical origin for this surprisingly large effect of a grain-boundary dislocation is that the superconducting properties of $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ are very sensitive to strain. Superconductivity is produced in this normally antiferromagnetic insulator by the addition of charge to the CuO_2 planes¹⁷. Disruption of the charge reservoir layers, for example by disordering the Cu–O chain layers, will affect superconductivity adversely. A strain of only 1% along the a or b axis is required to prevent the transformation from a tetragonal to an orthorhombic structure during processing, interfering with the order in the Cu–O chain layers and leading to significant volumes of non-

superconducting material surrounding each dislocation core. Independent support for this 1% criterion comes from particle irradiation experiments in which oxygen is displaced with no overall change in oxygen stoichiometry^{18,19}. Extrapolating the data to one oxygen displacement per unit cell, which would destroy the orthorhombicity, predicts T_c to be depressed by ~ 100 K. Further indications of the importance of strain comes from observations on $\text{Nd}_{1.83}\text{Ce}_{0.17}\text{CuO}_x/\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ multilayers, in which 1% strains significantly affected the normal-state resistivity and the superconducting transition²⁰. We therefore adopt 1% strain as the criterion for defining r_m , and calculate the strain field around an edge dislocation array parallel to the tilt axis assuming that linear isotropic elasticity theory is applicable²¹. Figure 3 is a plot of the roughly elliptical area around the dislocations in which one of the strain components is calculated to be $\geq 1\%$. Reduced distances are used so that the solutions for 2° , 5° and 10° tilt boundaries can be displayed. It can be seen that for all three cases there is a 'conducting plate' of relatively undistorted crystal between dislocations. For the 10° boundary, however, the thickness of this conducting plate is less than the dimensions of the unit cell, and thus the structural order required for superconductivity is destroyed everywhere along the boundary. For misorientation angles less than 10° , the equations for the extent of the 1% strain predict, to first order, linear dependence on misorientation angle, consistent with the available experimental data. The strong coupling observed by Mannhart *et al.*³ in bicrystal films with less than 10° misorientation, where the magnetic-field dependence of the grain boundary critical current density was similar to that of a single crystal, is entirely consistent with our observation that for tilts of up to 10° there are regions of the boundary that are largely unaffected by the dislocations. The grain-boundary dislocations effectively only reduce the area of the grain-boundary plane. Therefore, our microscopic characterization of the structure and composition of the boundaries, and our calculations of the dislocation strain field provide a consistent and physical explanation of the critical current behaviour of individual boundaries.

Beyond 10° misorientations, the boundary loses the required structural order for superconductivity, provided that it cannot relax to a lower-strain configuration, and becomes a weak link which is very sensitive to small applied magnetic fields³. Bulk-scale bicrystal results¹¹ and observations of special high-angle misorientations where a coincidence site lattice is produced²³ indicate, however, that there are grain boundaries that can undergo further relaxations to reduce the boundary energy; some of these may not behave as weak links. Unfortunately, the strain field at the boundary cannot be predicted solely by using geometrical factors that describe the boundary periodicity, but depends also on their specific atomic structure. The observation

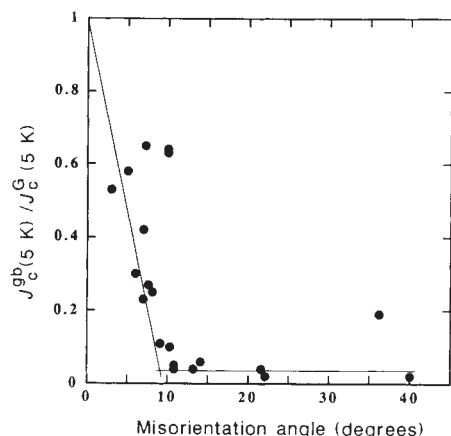


FIG. 2 Normalized critical current density as a function of misorientation⁴, plotted on a linear scale.

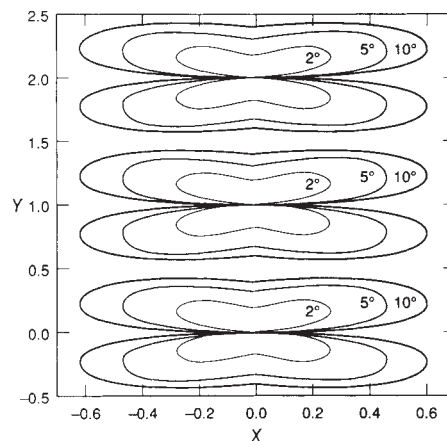


FIG. 3 Calculated strain field for an infinite array of edge dislocations at a 2° , 5° and 10° tilt boundary, showing the contour for a strain (ϵ_{xx} , the largest strain component) $\geq 1\%$, plotted using reduced distances $Y = y/D$ and $X = x/D$, where D is the dislocation spacing.

that a 14° boundary does not exhibit weak-link behaviour¹¹ does not necessarily contradict the strain model and, in fact, would seem to indicate that low-strain boundaries are not limited to high-coincidence orientations or, more probably, that the boundary contains segments of different structures, including a low-angle segment. The observation that 90° boundaries can sustain large critical currents^{11,12} is entirely consistent with our model, because the dislocation array for such boundaries will be widely spaced. The structure of these and other special boundaries is currently under investigation. As it seems that the strain field, rather than the dislocations or boundaries themselves, is responsible for the suppression of superconductivity, the implications of our results are not limited to grain boundaries. The surprisingly large effect of strain will influence the superconducting properties of many artificial structures, such as superlattices and thin-film devices, which involve strains of the order of 1%. □

Received 27 December 1990; accepted 19 March 1991.

1. Chaudhari, P. *et al.* *Phys. Rev. Lett.* **60**, 1653-1656 (1988).
2. Dimos, D., Chaudhari, P., Mannhart, J. & LeGoues, F. K. *Phys. Rev. Lett.* **61**, 219-222 (1988).
3. Mannhart, J., Chaudhari, P., Dimos, D., Tsuei, C. C. & McGuire, T. R. *Phys. Rev. Lett.* **61**, 2476-2479 (1988).
4. Dimos, D., Chaudhari, P. & Mannhart, J. *Phys. Rev. B* **41**, 4038-4049 (1990).
5. Camps, R. A. *et al.* *Nature* **329**, 229-232 (1987).
6. Nakahara, S. *et al.* *J. Cryst. Growth* **85**, 639-651 (1987).
7. Babcock, S. E. & Larbalestier, D. C. *Appl. Phys. Lett.* **55**, 393-395 (1989).
8. Kroeger, D. M. *et al.* *J. appl. Phys.* **64**, 331-335 (1988).
9. Pennycook, S. J. & Boatner, L. A. *Nature* **336**, 565-567 (1988).
10. Pennycook, S. J. & Jesson, D. E. *Phys. Rev. Lett.* **64**, 938-941 (1990).
11. Babcock, S. E., Cai, X. Y., Kaiser, D. L. & Larbalestier, D. C. *Nature* **347**, 167-169 (1990).
12. Clarke, D. R., Shaw, T. M. & Dimos, D. *J. Am. Ceram. Soc.* **72**, 1103-1113 (1989).
13. Campbell, A. M. *Supercond. Sci. Technol.* **2**, 287-293 (1989).
14. Chisholm, M. F. & Smith, D. A. *Phil. Mag.* **59**, 181-197 (1989).
15. Pennycook, S. J., Berger, S. D. & Culbertson, R. J. *J. Microsc.* **144**, 229-249 (1986).
16. Rothman, S. J., Routbort, J. L. & Baker, J. E. *Phys. Rev. B* **40**, 8852-8860 (1989).
17. Cava, R. J. *Science* **247**, 656-662 (1990).
18. Summers, G. P., Burke, E. A., Chrisey, D. B., Natasi, M. & Tesmer, J. R. *Appl. Phys. Lett.* **55**, 1469-1471 (1989).
19. Marwick, A. D., Guarneri, C. R. & Manoyan, J. M. *Appl. Phys. Lett.* **53**, 2713-2714 (1988).
20. Gupta, A. *et al.* *Phys. Rev. Lett.* **64**, 3191-3194 (1990).
21. Cottrell, A. H. *Dislocations and Plastic Flow in Crystals* (Clarendon Press, Oxford, 1953).
22. Chan, S.-W. *et al.* in *High T_c Superconducting Thin Films* Am. Inst. Phys. Conf. Proc. No. 200 (ed. Stockbaur, R.) 172-189 (American Institute of Physics, New York, 1990).
23. Smith, D. A., Chisholm, M. F. & Clabes, J. *Appl. Phys. Lett.* **53**, 2344-2345 (1988).

ACKNOWLEDGEMENTS. This research was sponsored by the Division of Materials Sciences, US Department of Energy with Martin Marietta Energy Systems, Inc.

Surface-mediated alignment of nematic liquid crystals with polarized laser light

Wayne M. Gibbons, Paul J. Shannon, Shao-Tang Sun & Brian J. Swetlin

Research Center, Hercules Incorporated, Wilmington, Delaware 19894, USA

THE control of molecular alignment in liquid-crystal phases at macroscopic scales has been investigated extensively because of its importance in optical or optoelectronic applications, such as liquid-crystal displays¹. It is well established that liquid crystals can be aligned by an applied electric field, a magnetic field, a shear-flow field, mechanical grooving of the substrate surface or stretching of liquid-crystal polymer thin films^{2,3}. Here we report a new mechanism for liquid-crystal alignment that uses polarized laser light. We find that nematic liquid crystals in an illuminated region become oriented perpendicular to the direction of the electric-field polarization of the laser and remain aligned in the absence of the laser radiation. The liquid crystals can be reoriented again by subsequent illumination. This technique might have applications for large-area displays, optical memories, binary optics, adaptive optics and molecular micro-assembly.

We spin-coated a glass substrate with a thin layer of a silicone polyimide copolymer doped with a diazodiamine dye at a dye:polyimide weight ratio of 1:2. The chemical structure and absorption spectrum of the dye are shown in Fig. 1. We assembled a cell using this coated glass substrate as the top plate and a second glass substrate, coated only with polyimide, as the bottom plate, with a spacing of $11\ \mu\text{m}$ between them. Both plates were rubbed with a cloth before assembly, the rubbing directions being matched on assembly. The cell was then filled with the nematic liquid crystal ZLI-1982 (EM Chemicals, Hawthorne, New York) at room temperature. This is a eutectic mixture of phenylcyclohexane liquid crystal molecules. By using a polarization microscope, we observed the direction of the nematic phase to line up with the rubbing direction of the cell.

The cell was illuminated with a polarized argon ion laser (514.5 nm), with the direction of laser polarization parallel to the rubbing axis (Fig. 2). Within the illuminated region, the molecules of ZLI-1982 then assumed a twisted nematic structure. The molecules adjacent to the illuminated, dye-doped surface became oriented perpendicular both to their original direction and to the direction of the laser polarization, whereas molecules adjacent to the undoped polyimide surface remained aligned parallel to the rubbing direction. The image written in this way can therefore be seen with a pair of polarizers (Fig. 3). Similar control of liquid-crystal alignment was achieved by illuminating

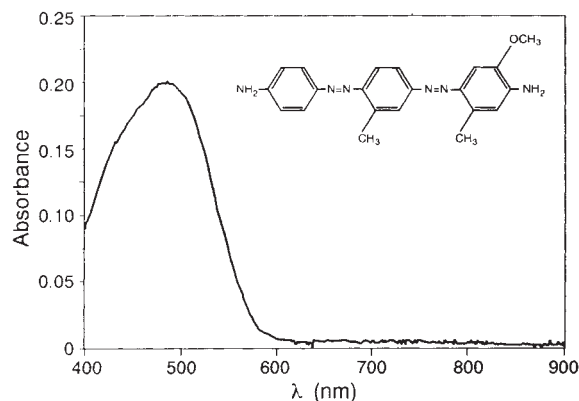


FIG. 1 The chemical structure and absorption spectrum of the dye dopant in the polyimide surface layer.

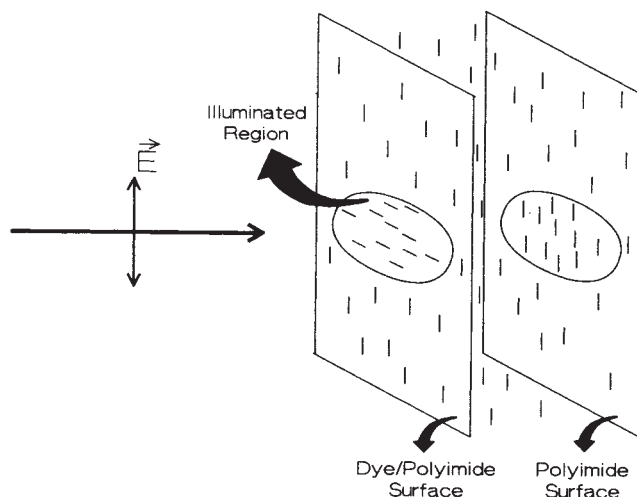


FIG. 2 The geometry of the illuminated liquid-crystal cell. The glass substrates of the cell are not shown for clarity. The rods represent the liquid-crystal orientation near the substrates before and after illumination.

the dye-doped polyimide substrate before filling with the liquid-crystal medium. A liquid-crystal 'grating' made in this way is illustrated in Fig. 4. The liquid-crystal alignment induced by laser irradiation is stable and can be erased or re-written by altering the direction of the incident electric-field polarization of the laser beam.

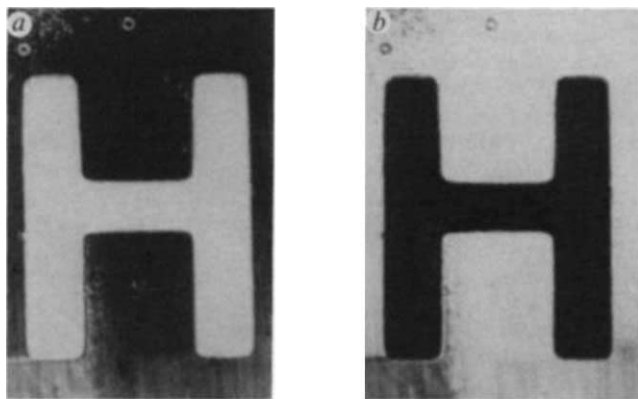


FIG. 3 A laser-aligned letter H in the liquid-crystal cell as viewed with a polarization microscope for *a*, parallel polarizers, and *b*, crossed polarizers. A mask in the shape of the letter H was placed in front of the liquid-crystal-filled cell, which was then illuminated with a 514.5-nm polarized laser at 9 W cm^{-2} for 1 min. The liquid-crystal alignment inside the letter H was along the initial alignment direction induced by rubbing the surfaces with a cloth. The liquid crystals within the illuminated background assumed a twisted nematic structure which caused a 90° rotation of the incident light polarization. The image has a height and width of 1.5 mm and 1.2 mm, respectively.

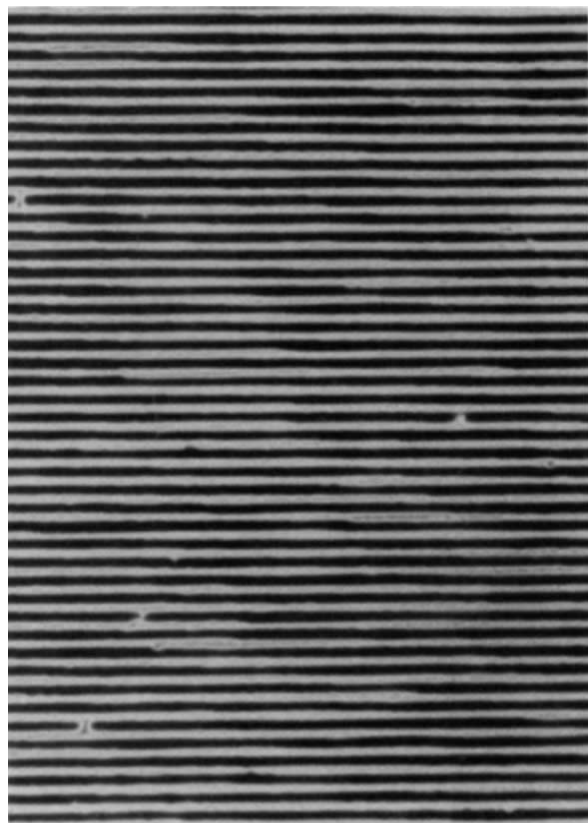


FIG. 4 A laser-aligned 10- μm grating in the liquid-crystal cell, as viewed with a polarization microscope with crossed polarizers. The cell was illuminated with a two-beam, plane-wave interference pattern with the angle between the plane waves chosen to give a fringe periodicity of 10 μm . The average incident intensity of the interfering beams onto the cell was 4 W cm^{-2} . The cell was exposed for 10 min. In this case, the cell was illuminated before filling with the liquid crystal.

An interesting aspect of our finding is that liquid crystals with a positive dielectric anisotropy are aligned perpendicular to the electric field of the laser beam. The 'memory' effect (alignment after irradiation of the dye-doped substrate) indicates that the laser-induced alignment is surface-mediated: the effect of irradiation of the dye-doped surface is apparently being 'read out' by the liquid crystal. Light-induced dye orientation and birefringence phenomena in a polymer matrix have been reported previously^{4,5}. It is plausible that the orientation of liquid crystals directly reflects the state of dye orientation at the liquid-crystal/polymer interface. The mechanism of coupling between liquid-crystal molecule and dye-doped polymer is unclear at present. Our preliminary results have shown that the kinetics of laser-induced orientation is a function of dye structure, dye concentration, liquid-crystal type, incident energy density and polymer chemistry.

The possibility of controlling the orientation of liquid crystals at a spatial resolution of the order of micrometres or less suggests a number of potentially useful applications. Liquid crystals are birefringent, and switchable by either electric or magnetic fields. Diffractive optical elements can thus be made, such as phase gratings and binary holograms. Coupled with proper electrode design, optical devices useful for adaptive optics can be envisaged. Our experiments also demonstrated the use of the liquid crystal as a read-out device for the surface state. The ordering of the surface by the laser beam may have uses in molecular micro-assembly. □

Received 8 February; accepted 7 March 1991.

1. Cognard, J. *Molec. Cryst. Liq. Cryst.* **51**, 1-77 (1982).
2. De Gennes, P. G. *The Physics of Liquid Crystals* (Clarendon, Oxford, 1974).
3. Geary, J. M., Goodby, J. W., Kmetz, A. R. & Patel, J. S. *J. appl. Phys.* **62**, 4100-4108 (1987).
4. Todorov, T., Nikolova, L. & Tomova, N. *Appl. Opt.* **23**, 4309-4312 (1984).
5. Ebralidze, T. D. & Mumladze, A. N. *Appl. Opt.* **29**, 446-447 (1990).

ACKNOWLEDGEMENTS. We thank S. T. Schnelle for technical support.

Insignificant isotropic component in the moment tensor of deep earthquakes

Hitoshi Kawakatsu

Geological Survey of Japan, 1-1-3 Higashi, Tsukuba, Ibaraki 305 Japan

THE mechanism responsible for deep-focus earthquakes, which occur at depths of 300-700 km in subducting slabs, has been a long-standing problem in geophysics. Unlike shallow earthquakes, deep earthquakes cannot be attributed to frictional instabilities across a fault plane, because of high frictional resistance to sliding at depth. A volumetric change associated with a phase transition, expected to occur at depth¹⁻³, is often invoked as the physical mechanism; if so, the resulting source mechanism should contain a major isotropic component. Although many researchers have attempted to observe such an isotropic component⁴⁻¹⁰, no one has yet convincingly proved or disproved its presence. There exists a portion of the seismogram which is well suited to resolve the isotropic component of deep earthquakes but which has not been analysed by previous workers. Here I use this component in a systematic analysis of 19 large deep earthquakes, and show that no significant isotropic component (<10% of the seismic moment) exists. A sudden implosive phase change can thus be ruled out as the primary physical mechanism for deep earthquakes.

A point-source earthquake mechanism is most generally expressed by a symmetric second-order tensor called the seismic moment tensor^{11,12} M_{ij} ($i, j = r, \theta, \phi$). The moment-tensor formalism includes both the isotropic source model and a double couple (two perpendicular force dipoles, with equal magnitudes

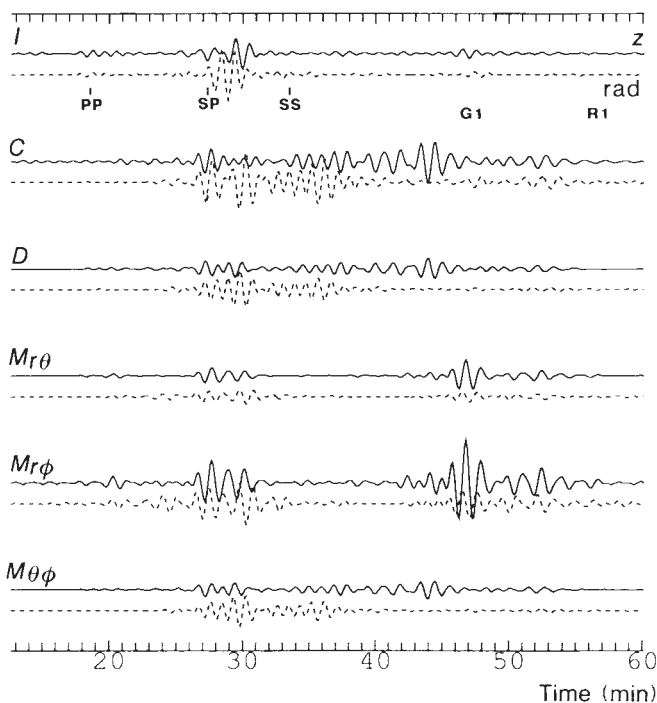


FIG. 1 Synthetic seismograms of long-period body waves for a deep earthquake. The earthquake is located at the hypocentre of the event south of Honshu and the station is BCAA (azimuth = 290°, $\Delta = 111^\circ$). The waveforms are calculated by a summation of normal modes and filtered between 10 and 22 mHz. For each pair of seismograms, the upper trace (solid line) is for the vertical component and the lower trace (dotted line) is for the radial component. Six pairs of seismograms are shown, one for each of six elements of the moment tensor. The diagonal components of a moment tensor are re-defined as $I = (M_{rr} + M_{\theta\theta} + M_{\phi\phi})/3$, $C = (M_{\theta\theta} + M_{\phi\phi} - 2M_{rr})/3$, $D = (M_{\theta\theta} - M_{\phi\phi})/2$. Note that the body-wave waveforms for the isotropic component and C -component are quite different, but this is not the case for long-period surface waves. Expected arrival times of some important phases are also indicated.

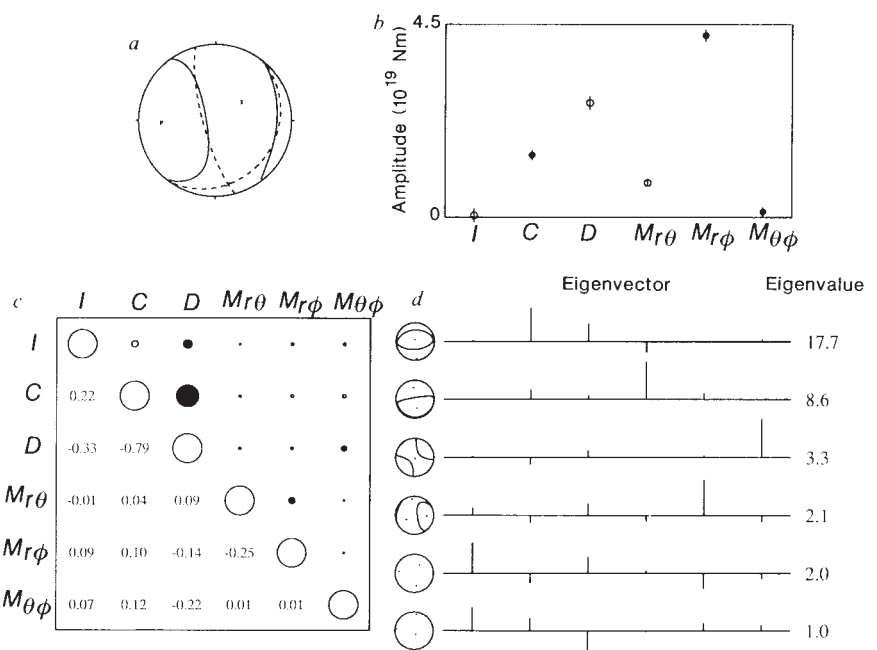
but opposite signs), which corresponds to a slip on a planar fault, as special cases. A double-couple model has four degrees of freedom to specify a fault plane, a slip direction and a slip size, whereas a moment tensor generally has six degrees of freedom. These two excess degrees of freedom are called non-double-couple components of the moment tensor. It is common to separate a moment tensor into two parts, the isotropic component $I = (M_{rr} + M_{\theta\theta} + M_{\phi\phi})/3$ and the deviatoric part $M_{ij} - I\delta_{ij}$. Any deviatoric non-double-couple moment tensor can be expressed as a summation of two double couples, although the decomposition is not unique.

It is difficult to observe the isotropic component because the long-period surface-wave waveforms or narrow-band low-frequency normal-mode data excited by an isotropic source are very similar to those excited by another independent (deviatoric)

component of a moment tensor, the vertical compensated linear vector dipole⁴ (CLVD) component $C = (M_{\theta\theta} + M_{\phi\phi} - 2M_{rr})/3$ (ref. 13). As most previous studies used only the above data, reported isotropic components, large or small, are subject to large uncertainty because of the presence of vertical CLVD components. Studies^{4,14} using amplitudes of a few body-wave phases are also not promising, because many deep earthquakes are known to have complex source time-functions¹⁵ and variable source mechanisms^{16,17}, and because the data coverage on the focal sphere is very limited. It is thus essential to analyse waveforms of many different body-wave phases to measure the isotropic component accurately.

The portion of the seismogram from the first P-wave arrival to just before the arrival of the first surface-wave train contains many different body-wave phases, including P, PcP, PP, PS and

FIG. 2 Example of CMT inversion for the south of Honshu earthquake (1 January 1984, depth = 383 km). *a*, Equal-area projection (lower hemisphere) of the CMT solution. The solid lines show the nodal lines of the full moment tensor, with the shaded area indicating the positive first motion region. The dotted lines indicate the nodal lines of the double-couple model that shares the same principal axes as the non-double-couple (full) moment tensor. Note the large non-double-couple component. *b*, Relative amplitude of each component of the moment tensor. The error bars indicate one standard deviation. Open and closed circles indicate positive and negative values, respectively. Note that the isotropic component is almost zero. *c*, Corresponding correlation matrix. The diameter of circle is proportional to the magnitude of the correlation coefficient. Here the correlation coefficient between I and C is 0.22, whereas that with surface-wave data is 0.80, showing the superior resolution of I -component using long-period body-wave waveforms. *d*, Eigensolutions of the normal equation matrix. The lengths of the stick marks on each row give the contribution of moment-tensor elements for each eigenvector. The corresponding relative eigenvalues are given on the right and corresponding focal mechanisms on the left. A focal mechanism (that is, an eigenvector) that has a larger eigenvalue is better resolved with the given data set. The dots on the focal spheres indicate locations of the principal axes. Note



that the eigenvectors that have significant isotropic components have the smallest eigenvalues.

others. Synthetic seismograms of such long-period (10–22 mHz) body waves for a deep earthquake indicate that the isotropic and the vertical CLVD components excite these waves quite differently (Fig. 1). It should therefore be possible to distinguish these two components by analysing long-period body waves; numerical experiments using these synthetic seismograms in fact prove that the isotropic component of a deep earthquake can be independently resolved by analysing long-period body-wave waveforms with sufficient station coverage (with stations every 30° both in longitudinal and azimuthal directions, the correlation between two components can be <0.05 for deep earthquakes). Similar experiments with long-period (~ 150 to ~ 300 s) surface waves show a strong correlation between the isotropic component and the vertical CLVD component. The vertical CLVD component is often quite large for deep earthquakes. If the lateral heterogeneity of the Earth is not included in a model, such strong correlation with a presence of a vertical CLVD component can bias the estimate of the isotropic component^{9,13}.

The centroid-moment-tensor (CMT) inversion method of Dziewonski *et al.*¹⁸, which simultaneously solves for a moment tensor and a spatio-temporal centroid, is well suited for inverting these long-period body-wave phases to obtain the isotropic component of the moment tensor. Figure 2 shows the result of the CMT inversion of the above data for a deep (~ 380 km) earthquake which occurred south of Honshu, Japan on 1 January 1984. This event is well suited to test the method, because it has

good station coverage and because it has a deviatoric non-double-couple moment tensor which is well documented^{17,19} and large, indicating a somewhat unusual source mechanism. The correlation matrix indicates that the isotropic component is well resolved; the isotropic component obtained is essentially zero. Figure 2d indicates that the eigenvectors that have significant isotropic components have the smallest eigenvalues. This suggests that it is still relatively difficult to observe the isotropic component even using long-period body-wave data. This is because the isotropic component is less efficient in exciting seismic waves in general, but the solution at least will not be biased by the presence of the vertical CLVD component.

To ascertain whether or not this result is general, we performed systematic body-wave CMT inversions for two sets of deep earthquakes. The first set consists of 9 events (including the earthquake on 1 January 1984) whose reported moment-tensor solutions²⁰ exhibit large deviatoric non-double-couple components. The second set consists of 10 events whose seismic moments are greater than 10^{19} N m. Figure 3 shows the moment-tensor solutions of the events in the first set. For all events that have large deviatoric non-double-couple components, the isotropic components are zero or negligible. Thus the large non-double-couple moment tensor observed for those events is entirely of deviatoric origin. For the event of 1 January 1984 it has been shown¹⁷ that the large non-double-couple component observed at low frequencies can be explained by the presence of multiple subevents with different double-couple focal mechanisms separated in time by a few seconds, and does not require any isotropic component.

Figure 4 summarizes the size distribution of the isotropic component for all 19 deep earthquakes studied. The isotropic component is always $\leq 10\%$ of the deviatoric seismic moment. Error bars covering two standard deviations always cut through the zero isotropic value. There seems to be no systematic difference between events with large non-double-couple components and those without. The sign of the isotropic component shows no systematic pattern, whereas an implosive phase change should always result in a negative isotropic component. These

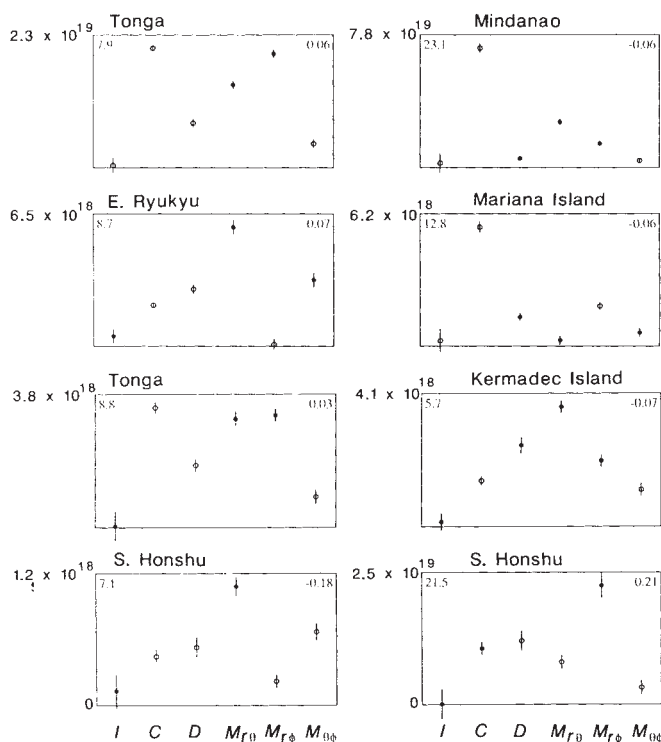


FIG. 3 Moment-tensor solutions for eight deep earthquakes which have a large deviatoric non-double-couple component (symbols as for Fig. 2b). The dates and depths of these earthquakes are: Tonga, 16 June 1986, 565 km; Mindanao, 5 March 1984, 630 km; east of Ryukyu, 4 July 1982, 554 km; Mariana Island, 4 January 1982, 590 km; Tonga 7 October 1981, 627 km; Kermadec Island, 28 September 1981, 335 km; south of Honshu, 31 March 1980, 359 km; south of Honshu, 7 March 1978, 434 km. The correlation coefficient between the I and C components, and the ratio of the maximum to minimum eigenvalues are shown at the right-upper and left-upper corner of each box, respectively. The ratio gives the condition number of the normal equation matrix; the smaller the ratio, the more stable the solution. Note that the vertical CLVD component (C) is often quite large for these events, and this can bias estimates of the isotropic component using long-period surface wave or free oscillation data. For all events, the isotropic components are essentially zero or negligible.

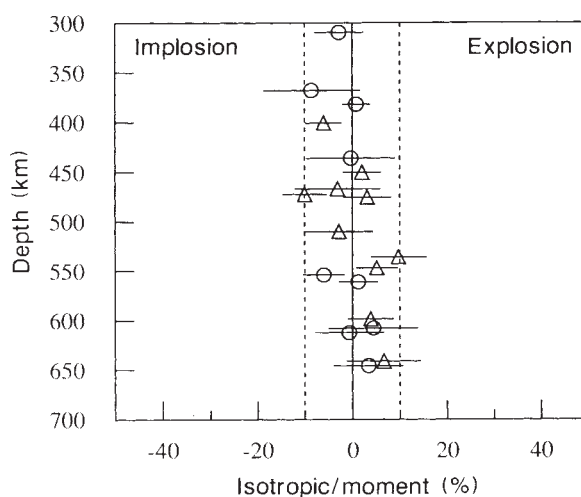


FIG. 4 Depth distribution of the isotropic components of all 19 deep earthquakes. The horizontal axis shows the size of the isotropic component as a percentage of the deviatoric seismic moment. The positive and negative values correspond to explosive and implosive isotropic sources, respectively. A deviatoric seismic moment is the average of the absolute values of the largest and smallest eigenvalues of a deviatoric moment tensor, $D_{ij} = M_{ij} - I \cdot \delta_{ij}$. The circles are for the nine events with large deviatoric non-double-couple components, and the triangles are for the ten events with seismic moments $> 10^{19}$ N m. The horizontal error bars indicate one standard deviation. The vertical broken lines indicate $\pm 10\%$ isotropic values and all points are within this range. There seems to be no systematic pattern and no difference between the two sets of deep earthquakes.

observations also hold for smaller deep events²¹. I therefore conclude that the isotropic component of deep earthquakes is <10% of the seismic moment in the period range 45–100 s, that its presence is statistically insignificant, and that a sudden implosive phase change can be ruled out as the primary physical mechanism for deep earthquakes. This conclusion is consistent with recent laboratory experiments on metastable phase transformations^{22,23}, which suggest 'transformation faulting' as the physical mechanism of deep earthquakes without requiring a significant isotropic component¹. □

Received 2 January; accepted 8 March 1991.

1. Bridgman P. W. *Am. J. Sci.* **A243**, 90–97 (1945).
2. Benioff, H. *Bull. seism. Soc. Am.* **53**, 893–903 (1963).
3. Evison, F. F. *Bull. seism. Soc. Am.* **57**, 9–25 (1967).
4. Randall, M. J. & Knopoff, L. *J. geophys. Res.* **75**, 4965–4976 (1970).
5. Dziewonski, A. M. & Gilbert, F. *Nature* **247**, 185–188 (1974).
6. Gilbert, F. & Dziewonski, A. M. *Phil. Trans. R. Soc. Lond.* **A278**, 187–269 (1975).
7. Silver, P. G. & Jordan, T. H. *Geophys. J. R. astr. Soc.* **70**, 755–787 (1982).
8. Riedesel, M. A. & Jordan, T. H. *Bull. seismol. Soc. Am.* **79**, 85–100 (1989).
9. Okal, E. A. & Geller, R. J. *Phys. Earth planet. Inter.* **18**, 176–196 (1979).
10. Vasco, D. W. & Johnson, L. R. *Geophys. J.* **97**, 1–18 (1989).
11. Gilbert, F. *Geophys. J. R. astr. Soc.* **22**, 223–226 (1970).
12. Backus, G. & Mulcahy, M. *Geophys. J. R. astr. Soc.* **46**, 341–361 (1976).
13. Mendigurren, J. A. & Aki, K. *Geophys. J. R. astr. Soc.* **55**, 539–556 (1978).
14. Stimpson, I. D. & Pearce, R. G. *Phys. Earth planet. Inter.* **47**, 107–124 (1987).
15. Fukao, Y. & Kikuchi, M. *Tectonophysics* **144**, 249–269 (1987).
16. Strelitz, R. A. *Phys. Earth planet. Inter.* **21**, 83–96 (1980).
17. Kuge, K. & Kawakatsu, H. *Geophys. Res. Lett.* **17**, 227–230 (1990).
18. Dziewonski, A. M., Chou, T. A. & Woodhouse, J. H. *J. geophys. Res.* **86**, 2825–2852 (1981).
19. Ekström, G., Dziewonski, A. M. & Stein, J. M. *Geophys. Res. Lett.* **13**, 173–176 (1986).
20. Dziewonski, A. M., Franzen, J. E. & Woodhouse, J. H. *Phys. Earth planet. Inter.* **34**, 209–219 (1984).
21. Kuge, K. thesis, Univ. Tokyo (1991).
22. Kirby, S. H. *J. geophys. Res.* **92**, 13789–13800 (1987).
23. Green, H. W. & Burnley, P. C. *Nature* **341**, 733–737 (1989).

ACKNOWLEDGEMENTS. I thank B. Geller, G. Ekström and H. Houston for useful comments.

Lower-mantle viscosity constrained by seismicity around deglaciated regions

G. Spada*, D. A. Yuen†, R. Sabadini* & E. Boschi*

* Istituto di Geofisica, Dipartimento di Fisica, Università di Bologna, 40127 Bologna, Italy

† Department of Geology and Geophysics and Minnesota Supercomputer Institute, University of Minnesota, Minneapolis, Minnesota 55415, USA

KNOWLEDGE of the viscosity structure of the Earth's mantle is important for constraining models of mantle convection and isostatic rebound. Here we show that seismicity around the margins of deglaciated areas provides a constraint on the viscosity of the lower mantle, in addition to those previously proposed^{1,2}. Calculations using a spherical, viscoelastic Earth model show that the present-day magnitude of the stress fields induced in the lithosphere beneath the (now-disappeared) Laurentide and Fennoscandian ice sheets is very sensitive to the value of the lower-mantle viscosity. Stresses of ~100 bar, sufficient to cause seismicity, can still remain in the lithosphere for lower-mantle viscosities greater than ~10²² Pa s; for lower-mantle viscosities of ~10²¹ Pa s, only a few tens of bars of stress persist in the lithosphere today. This influence of lower-mantle viscosity on the state of stress in the lithosphere also has implications for the migration of stress from earthquakes, and hence for earthquake recurrence times.

Estimates of mantle viscosity have been based on postglacial rebound¹ and geoid anomalies². A stratified mantle with a lower-mantle viscosity of ~10²² Pa s is favoured by geoid anomalies², whereas from global sea-level changes³ a lower-mantle viscosity has been obtained that is slightly higher than the upper-mantle viscosity of 10²¹ Pa s. Lower estimates for the upper-mantle viscosity of 3–5 × 10²⁰ Pa s can be derived from sea-level curves

in northwestern Europe⁴. The issue of mantle viscosity structure is by no means settled because of the intrinsic differences in the various data sets.

The seismic activity at passive continental margins in eastern Canada and Fennoscandia has been attributed to the melting of the Pleistocene ice sheets^{5–7}. Estimates of stress fields⁸ produced by removal of ice loads were based on the elastic plate flexure model and showed that considerable stresses (~10² bar) could be produced on unloading. Stresses from glacial melting were also computed using a viscous rheology in a spherical model with both uniform and stratified viscosity profiles. For a uniform mantle viscosity⁹, stresses of a few bars were found at the base of the lithosphere, whereas for individual long-wavelength harmonics¹⁰ in the stratified models, much higher stresses (~10² bar) still remained at the top of the mantle, and the lower-mantle viscosity was ~10²² Pa s. For the timescales considered here (~10⁴ yr), mantle deformation processes are better described by viscoelastic rheology. A more realistic model is needed to investigate the problem of contemporary stress fields induced by deglaciation; this model should incorporate viscoelastic relaxation, as the consequences of stress fields in former deglaciated regions can affect the assessment of earthquake hazards¹¹.

We have employed a five-layer model¹², consisting of a purely elastic lithosphere, a three-layer viscoelastic mantle with a Maxwell rheology and an inviscid core, to provide a realistic description of the evolution of stress caused by deglaciation and to account for the presence of a high-viscosity transition zone¹³ in the mantle. The shear modulus and density for each of the layers is taken from seismic estimates. The ratio of lower- to upper-mantle viscosity is denoted by B ; in models with a high-viscosity transition layer, there is another parameter C , representing the ratio of the transition-zone viscosity to the upper-mantle viscosity.

Green's functions can be constructed analytically for the stress fields from the elastic and viscoelastic contributions to the displacement fields. For the axisymmetric spherical model, the viscoelastic stress-tensor components in the Laplace-transformed domain are:

$$\sigma_{rr}(r, \theta; s) = \Pi + 2\mu(s) \sum_n^K \dot{U}_n P_n \quad (1)$$

$$\sigma_{r\theta}(r, \theta; s) = \mu(s) \sum_n^K \left[\left(-\frac{V_n}{r} + \dot{V}_n \right) P_n - \frac{U_n}{r} \partial_\theta P_n \right] \quad (2)$$

$$\sigma_{\theta\theta}(r, \theta; s) = \sigma_{rr} + 2\mu(s) \sum_n^K \left[\left(\frac{U_n}{r} - \dot{U}_n - \frac{n(n+1)V_n}{r} \right) P_n - \frac{\cot \theta}{r} V_n \partial_\theta P_n \right] \quad (3)$$

$$\sigma_{\phi\phi}(r, \theta; s) = 2\sigma_{rr} - \sigma_{\theta\theta} + 4\mu(s) \times \sum_n^K \left(\frac{U_n}{r} - \dot{U}_n - \frac{n(n+1)V_n}{2r} \right) P_n \quad (4)$$

where r and θ are respectively the radius and the co-latitude coordinate, ϕ is the longitude, s is the Laplace-transformed variable, n is the angular order of the Legendre function $P_n(\cos \theta)$, σ_{ij} are the elements of the stress tensor in an axisymmetric spherical model, Π denotes the isotropic components, and the scalar functions U and V represent respectively the radial and tangential displacements. The dot denotes differentiation with respect to r . The index K , here taken to be 100, represents the highest angular order used in summing the contributions from different wavelengths excited by the disintegrating ice sheet. These stress fields are time-dependent because the transformed shear modulus $\mu(s)$, U_n and V_n are all functions of the Laplace-transformed variable s . A linear Maxwell

reology is used in $\mu(s)$ with an intrinsic shear modulus μ_0 and long-term viscosity ν_0 for each of the layers¹⁴. This rheology has been commonly used^{2,12} in studying post-glacial rebound and changes from elastic to viscous behaviour over intermediate times of $\sim 10^3$ yr.

We have determined the evolution of the stress fields after a sequence of 10 glacial cycles with a period of 10^5 yr. The history of the displacement and stress fields is calculated for the past 12,000 years. We have investigated the stress fields induced by both the Laurentide and Fennoscandian ice sheets: these are modelled as circular disks that had fixed angular radii of 15° and 8° , respectively, and a parabolic profile along a vertical cross-section with initial maximum heights of 3.5 and 2.5 km respectively. The shrinkage of the ice sheets is controlled by decreasing the central peaks linearly with time. The details and the spectral coefficients used for the load are given in ref. 14.

Stress fields induced by deglaciation are influenced by stratification of the viscosity, as shown in Fig. 1 for a sequence of vertical sections of the top 400 km of the mantle. Each panel portrays a half cross-section passing through the axis of symmetry of the model, which is at the left. In Fig. 1a, we show a sequence of snapshots of $\sigma_M - \sigma_m$, the difference between the maximum and minimum eigenvalues of the stress tensor, for viscosity contrasts $B = 1, 5, 10, 50$ and 100 . Each of the columns of the four quadrants denotes an individual physical model. The ice load is representative of the Laurentide ice sheet with an elastic lithosphere of 100 km thickness.

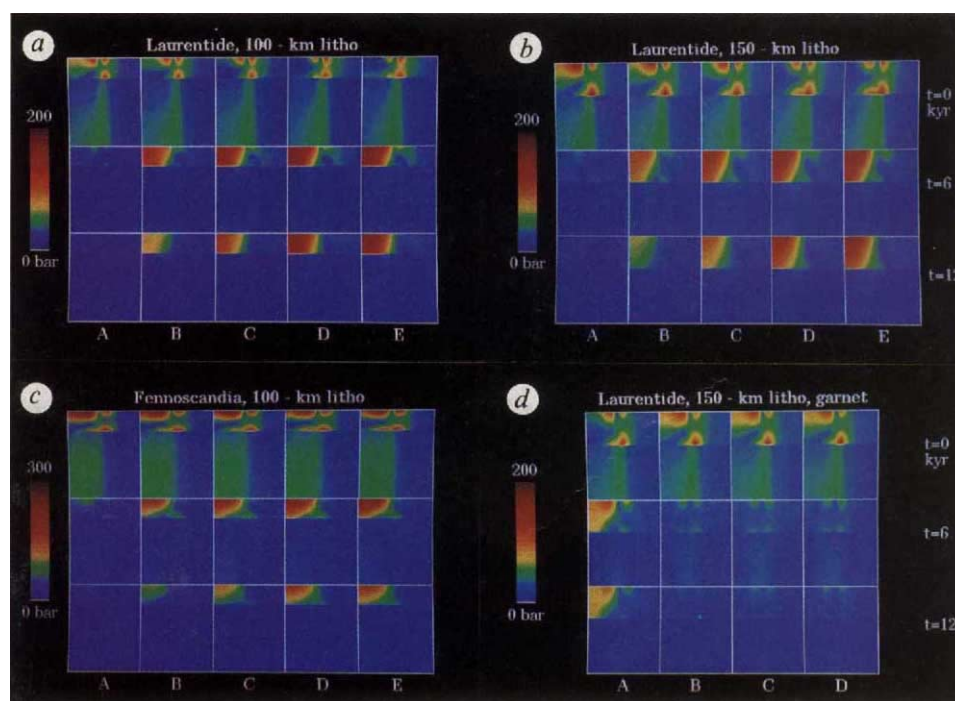
The stress fields at $t = 0$ (12,000 years ago) represented the cumulative effects from previous ice ages. There was a migration of stress from the mantle and a subsequent concentration of stress in the lithosphere (indicated by the red colour) at the edge of the ice sheet (0.4 the width of the panel), which spread downward into the mantle (green colour). After 6,000 years there was a build-up of stress in the lithosphere beneath the load for lower-mantle viscosity $\geq 5 \times 10^{21}$ Pa s or $B \geq 5$. Today, the lithospheric stress fields have decreased somewhat, as shown by the disappearance of the green colour at the edge of the former ice load. For a lower-mantle viscosity of 10^{21} Pa s, a

value commonly inferred¹ from post-glacial uplift, the stresses induced by deglaciation in the lithosphere are very small (blue colour), around 20 bar. Thus the amount of residual stress from the last ice age increases markedly for lower-mantle viscosity between 10^{21} and 10^{22} Pa s and flattens out if the mantle is highly stratified ($B > 50$). The thickness of the lithosphere can also influence the stress distribution: Fig. 1b shows the results for a lithosphere of thickness 150 km, characteristic of continental cratons¹⁵ beneath the Canadian ice sheet. A thicker lithosphere supports less residual stress, as can be seen by the lighter hues of red at $t = 12,000$ yr. For a lower-mantle viscosity of 10^{22} Pa s, a decrease in the stress levels of $\sim 20\%$ is found today. The trend of increasing stress with increasing lower-mantle viscosity is similar to that for the thinner lithosphere.

Stress fields are also affected by the lateral extent of the ice load. In Fig. 1c, we examine the effects of a smaller ice load such as that of the Fennoscandian ice sheet (model described above). Initially larger stresses, as shown by the green colour, are built up in the mantle and lithosphere (red colour) because the smaller size of the Fennoscandian ice sheet tends to focus the stress. Thereafter the stresses increase and then decrease to present-day values, which are comparable with those associated with the Laurentide melting for the same lithospheric thickness of 100 km (compare with Fig. 1a). Inspection of the bottom row shows that the effects of the lower-mantle viscosity on lithospheric stress fields are still evident today and are not influenced by the smaller size of the load.

Recent laboratory investigations¹³ and Monte Carlo inversion of geophysical surface signatures¹⁶ have indicated that the transition zone of the mantle at a depth of 400–670 km may be different from both the upper and lower mantles. Figure 1d shows the results of such a high-viscosity layer on the stress fields produced by melting of the Laurentide ice sheet over a lithosphere 150 km thick. We compare the case with $B = 10$ and high viscosity, $C = 50$ (panel A), with the cases where $B = 1$ and $C = 10, 50$ and 100 (panels B, C and D). The dominant influence of the lower-mantle viscosity over the high-viscosity layer is shown by the large stresses (yellow–orange) remaining in the lithosphere.

FIG. 1 a, Evolution of the $\sigma_M - \sigma_m$ field for the Laurentide ice sheet. σ_M and σ_m are the maximum and minimum principal stress values, respectively. The lithosphere thickness is 100 km. The panels depicting $t = 0$ do not include the elastic contributions from the instantaneous melting but reflect the contributions from the previous ten ice-age cycles with a period of 10^5 yr. Model A portrays an isoviscous mantle with a viscosity of 10^{21} Pa s ($B = 1$). Columns B, C, D and E have lower-mantle viscosities of 5×10^{21} , 1×10^{22} , 5×10^{22} and 10^{23} Pa s respectively ($B = 5, 10, 50$ and 100). The depth of each panel is 400 km, and the width is 2.5 times the angular radius of the ice-sheet, here taken to be 15° . The colour palette has 25 evenly spaced intervals and is linear. Maximum value of $\sigma_M - \sigma_m$ is < 200 bar (deep red). b, As in a, with a 150-km-thick lithosphere. c, Evolution of $\sigma_M - \sigma_m$ field for Fennoscandian ice sheet with a 100-km-thick lithosphere. The angular radius of the ice sheet is 8° . Same viscosity models ($B = 1, 5, 10, 50$ and 100) as in a. Scales of the colour palette go up to 300 bar. d, Evolution of $\sigma_M - \sigma_m$ for Laurentide ice load with a high-viscosity transition zone. The lithosphere thickness is 150 km. Model A has an upper-mantle viscosity of 10^{21} Pa s, a transition-zone viscosity of 5×10^{22} Pa s ($C = 50$) and a lower-mantle viscosity of 10^{22} Pa s



($B = 10$). Models B, C and D all have upper- and lower-mantle viscosities of 10^{21} Pa s and transition-zone viscosities of 10^{22} , 5×10^{22} and 10^{23} Pa s ($C = 10, 50$ and 100). Maximum $\sigma_M - \sigma_m$ (deep red) is 200 bar.

On the other hand, even for an extremely-high-viscosity transition zone ($C = 100$, panel D) the additional stress produced in the lithosphere is small, as shown by the faint blue colour.

We have quantified the stress induced in the lithosphere by calculating the volumetric average $\bar{\sigma}$ of $\sigma_M - \sigma_m$ over the thickness of the lithosphere and out to a horizontal distance of 1.5 times the radius of the ice sheet. The temporal evolution of $\bar{\sigma}(t)$ for the various models presented in Fig. 1 is shown in Fig. 2. The lower-mantle viscosity exerts the greatest influence on the evolution of the lithospheric stress. We observe a non-monotonic behaviour of $\bar{\sigma}$ as a function of time for lower-mantle viscosities $\geq 2 \times 10^{21}$ Pa s. This is due to the delayed relaxation of stress in the lower mantle, which was not considered previously. For these viscosities, lithospheric stresses decay with characteristic timescales much longer than those associated with a lower-mantle viscosity of $\sim 10^{21}$ Pa s. This has important implications for questions of seismic risk in regions that lay beneath large ice sheets in the late Pleistocene, such as eastern Canada and Fennoscandia.

The quantity $f = \sigma_{\theta\theta} - \sigma_{rr}$ yields important information about the stress state of the lithosphere. Negative and positive values of f represent tensional and compressional states, respectively. The stress pattern of f changes when the lower-mantle viscosity is increased to 5×10^{21} Pa s. It is characterized by tension in the lithosphere below the former ice sheets and by compression beneath the surrounding unglaciated region. This pattern agrees with seismic focal mechanisms⁷.

We have demonstrated that the stress fields induced by deglaciation 12,000 years ago are strongly influenced by the lower-mantle viscosity. For a highly viscous lower mantle (viscosity $\geq 10^{22}$ Pa s), an ice load characteristic of the late Pleistocene will give lithospheric stresses of 100–200 bar. These stresses could be large enough to induce seismic activity on faults that originated during previous tectonic episodes. Earthquake

activity along passive margins of eastern Canada and Fennoscandia is more consistent with a highly stratified viscous mantle.

Spatial and temporal changes of the stress fields in the mantle are also induced by abrupt motions along faults in earthquakes. Conventional viscoelastic models^{17,18} for the stress migration from earthquakes are based on the assumption of lithosphere–asthenosphere coupling^{18,19} and do not consider the effects of stratification of mantle viscosity at deeper levels. From our time-dependent calculations for stress redistribution after deglaciation, we propose that for large earthquakes, stress transfer by the stress-diffusion mechanism could have a long time-scale ($\sim 10^4$ yr), because of increased viscosity in the lower mantle. □

Received 15 October 1990; accepted 15 March 1991.

1. Cathles, L. M. III *The Viscosity of the Earth's Mantle* (Princeton University Press, 1975).
2. Richards, M. A. & Hager, B. H. *J. geophys. Res.* **89**, 5987–6002 (1984).
3. Peltier, W. R. & Tushingham, A. M. *Science* **244**, 806–810 (1989).
4. Lambeck, K., Johnston, P. & Nakada, M. *Geophys. J.* **103**, 451–468 (1990).
5. Gutenberg, B. & Richter, C. F. *Seismicity of the Earth and Associated Phenomena* (Princeton University Press, 1949).
6. Hasegawa, H. S. & Hermann, R. B. in *Earthquakes at North-Atlantic Passive Margins: Neotectonics and Postglacial Rebound* (eds Gregersen, S. & Basham, P. W.) 547–562 (Kluwer, Dordrecht, 1989).
7. Stein, S., Sleep, N., Geller, R., Wang, S. & Kroeger, G. *Geophys. Res. Lett.* **5**, 537–540 (1979).
8. Bungum, H. in *Earthquakes at North-Atlantic Passive Margins: Neotectonics and Postglacial Rebound* (eds Gregersen, S. & Basham, P. W.) 501–519 (Kluwer, Dordrecht, 1989).
9. Hager, B. H. & O'Connell, R. J. *J. geophys. Res.* **8**, 1031–1048 (1979).
10. Wu, J. & Yuen, D. A. *Geophys. J.* **104**, 331–350 (1991).
11. Wahlström, R. in *Earthquakes at North-Atlantic Passive Margins: Neotectonics and Postglacial Rebound* (eds Gregersen, S. & Basham, P. W.) 467–482 (Kluwer, Dordrecht, 1989).
12. Spada, G., Yuen, D. A., Sabadini, R. C. A., Morin, P. J. & Gasperini, P. *Mathematica J.* **1**, 65–69 (1990).
13. Karato, S. *Phys. Earth planet. Inter.* **55**, 234–240 (1989).
14. Yuen, D. A., Sabadini, R. C. A., Gasperini, P. & Boschi, E. V. *J. geophys. Res.* **91**, 11420–11438 (1986).
15. Karner, G. D., Steckler, M. A. & Thorne, J. A. *Nature* **304**, 250–253 (1983).
16. Ricard, Y., Vigny, C. & Froidevaux, C. *J. geophys. Res.* **94**, 13739–13754 (1989).
17. Nur, A. & Mavko, G. *Science* **183**, 204–206 (1974).
18. Savage, J. C. & Prescott, W. H. *J. geophys. Res.* **83**, 3369–3376 (1978).
19. Bonafede, M., Dragoni, M. & Morelli, A. *J. geophys. Res.* **91**, 6396–6404 (1986).

ACKNOWLEDGEMENTS. We thank T. Larsen for information about Fennoscandia and M. Wooters and P. J. Morin for help in preparing the manuscript. This work was supported by the Italian Space Agency, and the LAGEOS-II program of NASA.

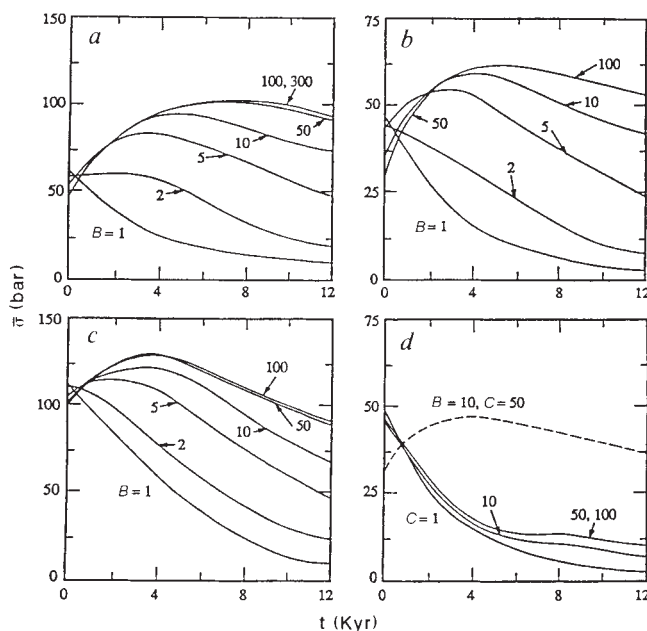


FIG. 2 History of volume-averaged $\sigma_M - \sigma_m$ for the various models. The average is taken over the thickness of the lithosphere and a horizontal distance 1.5 times the radius of the ice sheet. Panels a, b, c and d summarize the evolution of $\sigma_M - \sigma_m$ in Fig. 1. a and c are for a lithosphere thickness of 100 km, b and d for a thickness of 150 km. The dashed curve in d represents a model with both a high-viscosity transition zone ($C = 50$) and a high-viscosity lower mantle ($B = 10$). B and C are, respectively, the ratios of lower- to upper-mantle viscosity and of transition-zone viscosity to upper-mantle viscosity.

Low iron requirement for growth in oceanic phytoplankton

William G. Sunda*†, Dorothy G. Swift‡
& Susan A. Huntsman*

* National Marine Fisheries Service, Southeast Fisheries Center, Beaufort Laboratory, Beaufort, North Carolina 28516, USA

‡ Graduate School of Oceanography, University of Rhode Island, Narragansett, Rhode Island 02882, USA

† Present address: Oceanic Chemistry Program, Office of Naval Research, 800 N. Quincy Street, Arlington, Virginia 22217-5000, USA

DESPITE the controversy on the importance of iron in limiting phytoplankton growth and affecting air–sea exchange of CO_2 in the ocean^{1–4}, there is very little information on cellular iron requirements for growth. The few data available^{5,6} come from species isolated from coastal sea water where dissolved Fe levels are 10–1,000 times higher than those (≤ 0.1 nM) in the open ocean^{1,7}. Species from oceanic waters require much lower external Fe concentrations for growth than do comparable coastal species⁸. Here we report that an oceanic diatom was able to grow at a near maximum specific rate of about 1.0 per day at a cellular Fe:C ratio of 2 $\mu\text{mol}:\text{mol}$, about 25% of the amount needed for the same rate in a related estuarine species, and 2–20% of values previously used to estimate algal Fe requirements in sea water^{1,2}. These results have important implications concerning iron limitation of primary productivity in the ocean and cell biology of iron in oceanic algae.

We measured the effect of Fe concentration on cellular Fe:C ratio and growth rate in *Thalassiosira oceanica* (clone 13-1), isolated from the Sargasso Sea, and *T. pseudonana* (clone 3H)

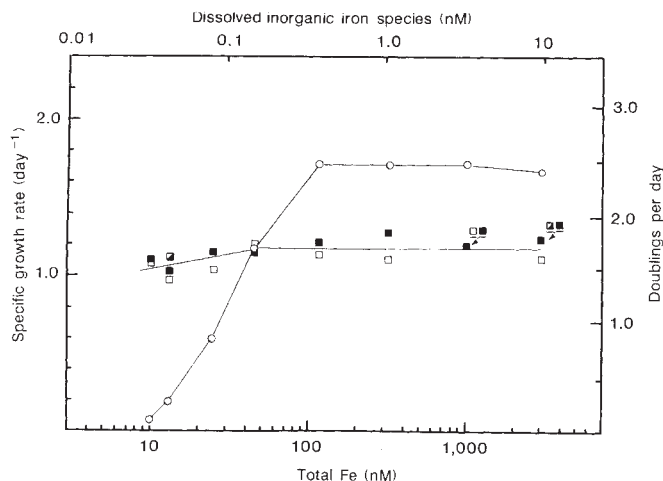


FIG. 1 Growth rate of clones 3H (○) and 13-1 (three successive experiments (■, ■, □)) as a function of total Fe and mean concentration of dissolved inorganic Fe species. Cells were preacclimated at low total Fe (10 nM for 13-1, 45 nM for 3H) for 7–19 days and grown at 20 °C and pH 8.10 ± 0.06 under 14:10 h light:dark cycle (600 $\mu\text{mol quanta m}^{-2} \text{s}^{-1}$) in filtered (0.4 μm) Gulf Stream sea water (stored in the dark at 7 °C for 1 year). Sea water contained added nutrients (35 $\mu\text{M NaNO}_3$, 1.5 $\mu\text{M Na}_2\text{HPO}_4$, 40 $\mu\text{M Na}_2\text{SiO}_3$, 0.1 $\mu\text{g l}^{-1}$ vitamin B_{12}) and a trace metal ion buffer system (0.1 mM EDTA, 100 nM CuCl_2 , 250 nM ZnCl_2 , 120 nM MnCl_2 , 100 nM CoCl_2). Total Fe was computed from the sum of the background concentration (10 nM as measured by atomic absorption spectrophotometry following organic extraction^{18,19}) plus Fe added with radiotracer and as unlabelled FeCl_3 . Log free-ion concentrations of Cu^{2+} , Zn^{2+} , Co^{2+} and Mn^{2+} , were computed from equilibrium theory^{11,20} and are -13.12, -10.42, -10.63, and -8.14, respectively. The mean dissolved inorganic Fe concentration $[\text{Fe}']$ was computed from conditional formation, dissociation and photodissociation rate constants for iron-EDTA in sea water⁵. Computed $[\text{Fe}']/[\text{total Fe}] = 0.00077$ in dark and 0.0053 in light giving weighted mean of 0.0034 for 14:10-h light:dark cycle.

from a eutrophic estuary⁹. Experimental procedures and conditions were similar to those in previous culture studies with ^{54}Mn (refs 10, 11). Algae were grown in filtered Gulf Stream sea water containing added nutrients, FeCl_3 radiolabelled with ^{55}Fe and an EDTA-trace metal ion buffer system that reduced available inorganic Fe to levels far below total Fe concentrations. Growth rate was measured from daily increases in total cell volume using a Coulter counter. Intracellular Fe was determined in exponentially growing cells by filtering them onto 3- μm -pore Nucleopore filters, rinsing briefly with a Ti(III)EDTA -citrate reducing solution¹² to dissolve iron hydroxides and desorb ferric ions bound to cell surfaces, and measuring the remaining particulate ^{55}Fe by liquid scintillation counting. In some instances the cells were rinsed with sea water instead of reducing solution to measure total (intracellular plus surface) cell Fe. Cell ^{55}Fe values were corrected with filter blanks using media without cells. The fraction of ^{55}Fe in the cells was multiplied by the total Fe concentration and divided by the measured volume of cells per litre to yield cell Fe concentrations. These were converted to Fe:C ratios using cell carbon:volume ratios of 22 and 15 mol C l^{-1} for clones 3H and 13-1 that we determined with standard ^{14}C techniques and Coulter counter measurements of cell volume. Carbon:volume ratios were unaffected by Fe concentration (data not shown).

Decreases in Fe concentration reduced growth rate more in

the neritic clone 3H than in the oceanic clone 13-1 (Fig. 1) in accordance with previous results⁸. Clone 3H grew 50% faster than clone 13-1 at high Fe concentrations; but at the lowest Fe level, its growth rate decreased to near zero while that of clone 13-1 was reduced by only about 15%. The ability of clone 13-1 to grow faster than clone 3H at low Fe levels was due almost entirely to a much lower cellular iron requirement for growth (Fig. 2), and not to a greater ability to accumulate iron (Fig. 3a). When intracellular Fe:C ratios were multiplied by specific growth rates, the resultant specific cellular Fe uptake rates were very similar for the two species at low growth-limiting Fe levels (Fig. 3b).

Oceanic algae might have been expected to have evolved higher affinity transport systems to acquire iron more effectively at low concentrations, as observed for other nutrients (for example Mn^{11} and nitrate¹³). But Hudson and Morel⁵ recently reported that Fe-uptake kinetics in two coastal phytoplankters approach the physical limits for diffusion of inorganic Fe species to the cell surface and for kinetics of Fe-ligand exchange at membrane transport sites. They predicted that oceanic algae could not have higher transport kinetics and that the only available means of adaptation to low-iron oceanic conditions would be a reduction in cell Fe requirement or size. A reduction in size does not apply in this case because clone 13-1 is larger than clone 3H; mean cell volumes were 139 ± 12 ($\pm \text{s.d.}$) and

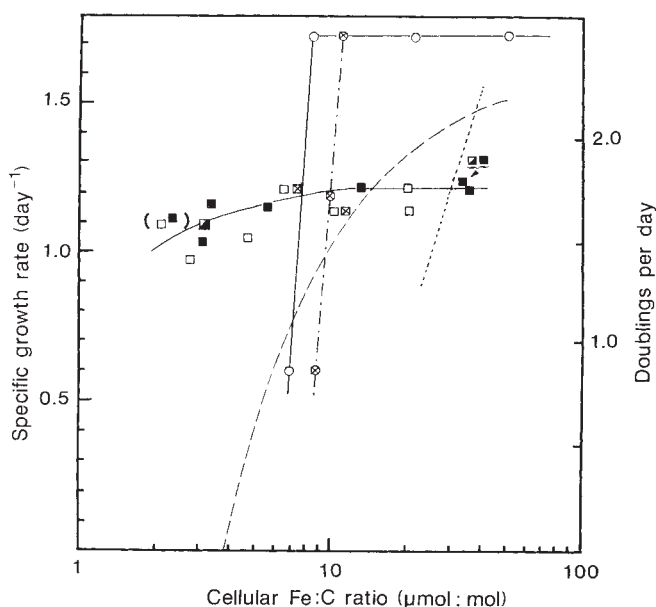


FIG. 2 Specific growth rate versus cell Fe:C ratio for a single 3H experiment based on intracellular (○) and total cell (⊗) Fe, for two 13-1 experiments based on intracellular Fe (■, ■) and a third based on intra- (□) and total (⊗) cell Fe. Ratios of intra- to total cell Fe were 0.89 ± 0.005 and 0.77 ± 0.1 ($\pm \text{range}$, $n = 2$) for 13-1 and 3H at Fe concentrations ≤ 110 nM where Fe(OH)_3 does not precipitate. Data points in parentheses (■, □) are estimates for cultures without added Fe or ^{55}Fe , extrapolated from specific growth rate and relationships among cell Fe uptake rate, specific growth rate and Fe concentration (Figs 1 and 3; refs 11, 20). Cells were inoculated at 0.3 $\mu\text{mol cell C l}^{-1}$ and grown for about 7 generations before measurement. Cell Fe levels were measured in narrow biomass range of 29–75 $\mu\text{mol cell C l}^{-1}$ to have sufficient cell mass for accurate measurement without having enough to appreciably affect dissolved Fe chemistry or culture pH (ref. 5). Dashed line is data for *T. weissflogii* from Harrison and Morel⁶ measured at 20 °C in continuous light normalized to cell C using our measured C:volume ratio of 5.6 mol l^{-1} . Dotted line is estimate of cell Fe:C growth requirement based on biochemical maximum use efficiency calculations of Raven¹⁴ at 20 °C and saturating light. His specific growth rates computed for continuous light were multiplied by 0.58 to adjust to our 14:10 h light:dark cycle.

$32 \pm 4.7 \mu\text{m}^3$, respectively. Our results confirm their predictions.

The amount of cellular Fe required for near maximum growth of clone 13-1 is much lower than minimum amounts thought necessary to meet the metabolic needs of plant cells¹⁴. It is $\leq 10\%$ of calculated amounts needed for growth based on Fe enzymatic requirements in photosynthesis, respiration and NO_3 reduction (Fig. 2). It is also 10–100 times less than values previously used to estimate algal growth requirements for Fe in sea water^{1,2}. These estimates were based on laboratory experiments with unacclimated cultures of the coastal species *T. weissflogii*¹⁵ and on amounts of particulate Fe in near-surface sea water after subtracting amounts leachable by weak acid and estimated to occur in aluminosilicate minerals². These latter values are uncertain as they do not correct for iron adsorbed to particles or present in iron oxides. On the basis of these values, it has been concluded that ratios of Fe: NO_3 in upwelling sea water were 10–100 times too low to meet the metabolic needs of phytoplankton and, therefore, that most of the Fe required for growth must be supplied from atmospheric deposition^{1,2}.

This conclusion needs to be reassessed in light of our findings. Dissolved Fe within the nutricline of the North Pacific is highly correlated with NO_3 , PO_4 and SiO_3 concentrations^{1,2}, suggesting that it is regulated by biological uptake and regeneration processes as occurs for major nutrients. For this to be true, however, the relative changes in dissolved Fe and major nutrients with depth should reflect the concentrations of these elements in phytoplankton, the major biomass reservoir. Linear regressions between dissolved Fe and NO_3 in the nutriclines of four North Pacific stations² yield slopes of 13.5, 15.0, 11.3 and $13.8 \mu\text{mol Fe}:\text{mol NO}_3$ (coefficient of variation, $r^2 = 0.98, 0.94, 0.96, 0.90$) which translate to a mean Fe:C ratio of $2.0 \pm 0.2 \mu\text{mol}:\text{mol}$, assuming a typical 6.6:1 C:N ratio in plankton¹⁶. This Fe:C ratio would support a specific growth rate of clone 13-1 of about

1.0 per day, suggesting that there could be enough dissolved Fe in deep water when it is advected to the surface to support low to moderate growth rates of oceanic species, assuming most of the Fe is biologically available.

Although the Fe:C ratio derived from dissolved Fe to major nutrient correlations will support growth of the oceanic species, it is too low for growth of coastal diatoms *T. pseudonana* and *T. weissflogii* (Fig. 2). Thus, for a hypothetical algal community composed of these three species, the coastal diatoms would be selected against under oceanic low-Fe conditions and the community would be dominated by the slower growing oceanic diatom. Increasing the Fe concentration might have little immediate effect on community growth rate as the oceanic species might be growing near its maximum; but after a period the rate would increase as the community became dominated by the high-Fe species with faster maximum rates. This would explain observations from algal growth experiments in surface water from the subarctic Pacific^{2,4} and Southern Ocean³ in which Fe additions caused little or no effect for the first 2–4 days, but stimulated growth and caused associated shifts in species dominance after this period.

Data from the experiments in the subarctic Pacific² support our findings of the low cellular Fe:C growth requirements of oceanic algae. There was some growth in all of the controls (with no Fe added) in these experiments, despite low Fe concentrations in the water. At station T-6, for example, suspended particulate matter increased by 0.814 mg kg^{-1} (6.2-fold), which apparently was due to phytoplankton growth as it was accompanied by a 7.4-fold increase in chlorophyll. Dissolved Fe was 0.08 nM , and if all of it were taken up by the growing phytoplankton, we compute a maximum cellular Fe:C ratio of $2.6 \mu\text{mol}:\text{mol}$ (based on a 0.45 cellular C:dry weight ratio¹⁴). We compute a similar value ($2.9 \mu\text{mol}:\text{mol}$) if we base our growth estimates on the amount of NO_3 taken up by the cells ($4.2 \mu\text{M}$) and a C:N ratio for plankton of 6.6 (ref. 16). These estimates agree well with our growth requirements for clone 13-1 and with the Fe:C ratio ($2.3 \mu\text{mol}:\text{mol}$) derived from the relative increases in dissolved Fe and NO_3 within the nutricline at this station as discussed above.

The results of our investigation, the covarying distributions of dissolved Fe and major nutrients, and the results of ship-board growth experiments all indicate that Fe is an important biologically controlling nutrient in the sea whose distribution is influenced by phytoplankton uptake and regeneration processes. Whether it is more important in controlling algal community growth rate and species composition than traditional major nutrients is unknown, and is currently being actively debated¹⁷. □

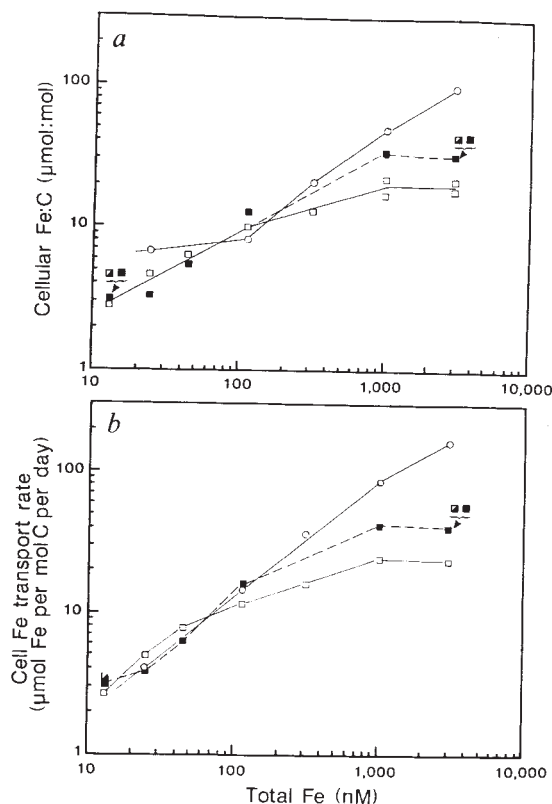


FIG. 3 a, Intracellular Fe:C ratios in clones 3H and 13-1 as functions of total Fe in medium. b, Specific Fe transport rates were computed by multiplying intracellular Fe:C ratios by specific growth rates. ○, 3H; □, ■ and ●, three experiments with 13-1.

Received 9 October 1990; accepted 7 March 1991.

- Martin, J. H. & Gordon, R. M. *Deep Sea Res.* **35**, 177–196 (1988).
- Martin, J. H., Gordon, R. M., Fitzwater, S. & Broenkow, W. W. *Deep Sea Res.* **36**, 649–680 (1989).
- Martin, J. H., Fitzwater, S. E. & Gordon, R. M. *Global Biogeochem. Cycles* (in the press).
- Coale, K. H. thesis, Univ. California, (1988).
- Hudson, R. J. M. & Morel, F. M. M. *Limnol. Oceanogr.* **35**, 1002–1020 (1990).
- Harrison, G. I. & Morel, F. M. M. *Limnol. Oceanogr.* **31**, 989–997 (1986).
- Gordon, R. M., Martin, J. H. & Knauer, J. *Nature* **299**, 611–612 (1982).
- Brand, L. E., Sunda, W. G. & Guillard, R. R. L. *Limnol. Oceanogr.* **28**, 1182–1198 (1983).
- Guillard, R. R. L. & Rhyther, J. H. *Canad. J. Microbiol.* **8**, 437–445 (1962).
- Sunda, W. G. & Huntsman, S. A. *Limnol. Oceanogr.* **28**, 924–934 (1983).
- Sunda, W. G. & Huntsman, S. A. *J. Phycol.* **22**, 259–270 (1986).
- Hudson, R. J. M. & Morel, F. M. M. *Limnol. Oceanogr.* **34**, 1113–1120 (1989).
- Carpenter, E. J. & Guillard, R. R. L. *Ecology* **52**, 183–185 (1971).
- Raven, J. A. *New Phytol.* **20**, 279–287 (1988).
- Anderson, M. A. & Morel, F. M. M. *Limnol. Oceanogr.* **27**, 789–813 (1982).
- Redfield, A. C., Ketchum, B. H. & Richards, F. A. in *The Sea* (ed. Hill, M. N.) 26–77 (Interscience, New York, 1963).
- Banase, K. *Limnol. Oceanogr.* **35**, 772–775 (1990).
- Bruland, K. W., Franks, R. P., Knauer, G. A. & Martin, J. H. *Analytica chim. Acta* **105**, 233–245 (1979).
- Landing, W. M. & Bruland, K. H. *Geochim. cosmochim. Acta* **51**, 29–43 (1987).
- Sunda, W. G. *Mar. Chem.* **14**, 365–378 (1984).
- Morel, F. M. M. *J. Phycol.* **23**, 137–150 (1986).

ACKNOWLEDGEMENTS. We thank R. Flegal and G. Smith for analysis of background iron in our growth medium and thank R. Hudson for calculations of dissolved inorganic iron species. This work was partially supported by the NSF (D.G.S.).

Sexual selection and the potential reproductive rates of males and females

T. H. Clutton-Brock & A. C. J. Vincent

Large Animal Research Group, Department of Zoology,
University of Cambridge, Cambridge CB3 0DT, UK

PRONOUNCED sex differences in mating competition are a prominent feature of many animal breeding systems. These differences are widely attributed to sex differences in parental investment^{1,2} which bias the ratio of sexually receptive females to males³ (the operational sex ratio), generating more intense competition between members of one sex, usually males³⁻⁵. Unfortunately, relative parental investment¹ is usually impossible to measure in species where both sexes invest in their offspring^{6,7} and there is currently no empirical basis for predicting the pattern of mating competition in these species. In contrast, the potential rate of reproduction by males and females (measured as the maximum number of independent offspring that parents can produce per unit time) is both more directly related to the operational sex ratio and more easily estimated in natural populations⁷. Here we show that among species where males care for the young, the sex with the higher potential reproductive rate competes more intensely for mates than the sex with the lower potential rate of reproduction.

In animals without parental care or where females are responsible for all care, the potential reproductive rate of males usually exceeds that of females. As a result, the operational sex ratio is biased towards males and males are the predominant competitors for mates^{1,7,14}, (except in a few cases where males contribute resources used in the production of zygotes⁸⁻¹¹). In contrast, in species where males are responsible for all parental care while females pay the costs of egg production (which include some teleost fishes¹⁴⁻¹⁶, anurans¹⁷, urodeles¹⁸, invertebrates^{19,20} and a few birds^{21,22}), the direction of mating competition differs between species. In some, females compete intensely for mates, males are choosy in selecting partners and females are brighter than males²³⁻²⁶. In others, males compete intensely for females, females are choosy in selecting partners, and males are brighter than females^{27,28}. An explanation of these differences could be that only in some of these species does the involvement of males in parental care depress their potential reproductive rate below that of females⁷.

To test whether differences in the direction of mating competition depend on which sex has the higher potential rate of reproduction, we extracted data on the maximal reproductive rates of males and females for 29 species where males were responsible for parental care, there was a clear sex difference in the intensity of mating competition, and data were available (Tables 1 and 2). With only two possible exceptions, males had potentially higher reproductive rates than females in all 'predominant male competitors' (Table 1). The most highly developed examples of predominant male competition combined with male parental care are found in fish and frogs where males can care for multiple clutches simultaneously or in quick succession²⁷⁻²⁹. For example, in the three-spined stickleback, *Gasterosteus aculeatus*, males can guard 10 or more clutches of eggs at a time and do so for about 2 weeks, whereas females can lay one clutch every 3-5 days²⁸. Consequently, the potential reproductive rate of males is higher than that of females, the operational sex ratio is male-biased and males compete intensely for mates.

By contrast, in all species that we identified as 'predominant female competitors', females were able to achieve higher rates of reproduction than males. The clearest examples occur in small polyandrous shorebirds, where the potential reproductive rate of males is low because incubation is prolonged and brood

TABLE 1 Relationship between reproductive rates and mating competition for species in which males are responsible for parental care

	Competition for mates more intense in males	Competition for mates more intense in females
<1	Fish <i>Cottus</i> (2 spp) <i>Oxylebius pictus</i> <i>Chromis notata</i> <i>Chrysiptera cyanea</i> <i>Badis badis</i> <i>Pimephales promelas</i> <i>Etheostoma olmstedii</i> <i>Gasterosteus aculeatus</i> <i>Forsterygion varium</i> Frogs <i>Alytes obstetricans</i> <i>Hyla rosenbergii</i> <i>Eleutherodactylus coqui</i>	
Female rate		
Male rate		
>1	Fish <i>Hippocampus</i> spp Birds ? <i>Rhea americana</i>	Fish <i>Apogon notatus</i> <i>Nerophis ophidion</i> <i>Syngnathus typhle</i> Birds <i>Actitis macularia</i> <i>Phalaropus</i> (2 spp) <i>Eudromias morinellus</i> <i>Jacana</i> (5 spp) <i>Rostrathula benghalensis</i> <i>Turnix sylvaticus</i>

Males compete more than females for access to mates in all but two of the species in which a male has a higher potential reproductive rate than a female (mainly ectotherms). Females are the more competitive sex (sex roles are 'reversed') in species where the potential reproductive rate of females exceeds that of males (primarily endotherms).

size is small²¹. For example, in the polyandrous spotted sandpiper *Actitis macularia*, where females compete intensely for mating partners, males do not raise more than one clutch of four eggs during the breeding season, whereas females can produce an egg a day and lay clutches for up to four different males in the course of the season^{23,30}. Predominant female competition also occurs in some fish where males carry eggs or young for lengthy periods and their reproductive rate is constrained by the number of eggs they can carry, including the pipefishes, *Nerophis ophidion* and *Syngnathus typhle*^{25,26,31,32}, and some cardinal fishes³³. Further examples can be expected in other animals where males bear eggs or young.

Both possible exceptions in Table 1 are instructive as they illustrate the need to calculate reproductive rates over different periods in different species. In the greater rhea, *Rhea americana*, males incubate broods of 20-30 eggs laid by several females and compete vigorously for mating access to female groups³⁶. The potential reproductive rate of females calculated over the entire breeding season may be higher than that of males. During the period of mating and brood production however, males can fertilize and accept eggs faster than females can lay them and the operational sex ratio is probably male-biased. In seahorses (*Hippocampus* spp.), the operational sex ratio is biased towards males despite a prolonged male gestation period because the reproductive rate of females is constrained by monogamous pair bonds and by limited periods of receptivity¹².

The potential rates of reproduction by males and females thus provide a basis for predicting the direction of mating competition in the two sexes and thus the direction of sexual selection. Several other factors, however, can bias the operational sex ratio and influence the relative intensity of mating competition. These include behavioural adaptations to competition in the sex with the potentially higher reproductive rate, such as precopulatory guarding of multiple mates and earlier eclosion, emergence or arrival times^{24,37-40}. Conversely, biases in the operational sex

TABLE 2 Maximum observed reproductive rates in species where males are responsible for parental care

Predominant male competitors						
	Male care duration	Clutch (C) and brood (B) size	Interclutch interval	Max. F/max. M rate of reproduction	Competing sex	Ref.
<i>Alytes obstetricans</i> (Discoglossidae: Midwife toad)	2-3 weeks	M can carry > 1 clutch at a time	FF breed 2-4 times per summer at ~ monthly intervals	<1	M	45-47
<i>Hyla rosenbergii</i> (Hylidae)	4 days	$B = 2,350$	23 days	<1 (0.25)	M	27
<i>Eleutherodactylus coqui</i> (Leptodactylidae)	17-26 days of care 8-9 month season	$B = \text{up to } 5 \times C$ at once $C = 16-43$	up to 6 clutches per season	<1	M	29
<i>Cottus hangiangensis</i> (Cottidae: river sculpin)		$B = 3.4-5.3 \times C$ up to $13 \times C$	2 per season	<1	M	49
<i>Cottus gobio</i> (Cottidae: river bullhead)	4 weeks	$B = 2.15 \times C$ $C = 75-200$	2 per season	<1	M	50
<i>Oxylebius pictus</i> (Hexagrammidae: greenling)	2.5-3.5 weeks 30 days between spawnings up to $7.5 \times C$ per season	$B = 0-10 \times C$ per cycle $B = 0-22 \times C$ per season $C = 1,500-5,000$	3 per season	<1	M	51, 52
<i>Chromis notata</i> (Pomacentridae: damselfish)	4-12 days (depends on temperature) 8.2-17.8 days between spawnings	$B = 1-4 \times C$ 2.8-4.6 nests per season $C = 10,000-27,500$	7.2-18.4 days	<1	M*	15, 53
<i>Chrysiptera cyanea</i> (Pomacentridae: damselfish)	4 days MM spawn continuously	$B = \text{up to } 12,255$ eggs per cycle $C = 900-2,500$ eggs per cycle	4 days	<1	M	54
<i>Badis badis</i> (Nandidae)	2 days egg/3-4 days larvae 7-8 days between spawnings	$B = 2-3 \times C$	4-7 days	<1	M	55, 56
<i>Pimephales promelas</i> (Cyprinidae)	4-8 days males spawn continuously for 3-5 weeks	$B = \text{max. } 6,000$ eggs per nest $C = 200-700$ eggs	3-4 days	<1	M	57, 58
<i>Etheostoma olmstedii</i> (Percidae: tessellated darter)	4 days	$B = \text{max. } 2,000$ eggs $C = 19-324$ (season $\lambda = 727$)	5-16 days ($\bar{x} = 7.6$)	<1	M*	15, 59
<i>Gasterosteus aculeatus</i> (Gasterosteidae: stickleback)	2 weeks at 21 °C	$B = \text{up to } 10 \times C$ $B \gg C = 51-150$	3-5 days	<1	M	28
<i>Forsterygion varium</i> (Tripterygiidae)	7-10 days spawn up to 15 times per season	$B = 20-7,080$ ($\bar{x} = 2,245$) eggs $C = 20-1,680$ ($\bar{x} = 796$) eggs	100% spawn once 31% spawn > once never > 5 times per season	<1 (0.3)	M	60
<i>Hippocampus fuscus</i> (Syngnathidae: Seahorse)	13-14 days	$C > B$	≤13-14 days	>1	M	12
<i>Rhea americana</i> (Greater Rhea)	incubation 36-37 days plus care one brood per season	$B = 26.5$ $C = 2-3$	up to 12 MM per year	eggs per season > 1 eggs per laying period < 1	M	34-36
Predominant female competitors						
	Male care duration (including incubation)	Clutch (C) and brood (B) size	Interclutch interval	Max. F/max. M rate of reproduction	Competing sex	Ref.
<i>Actitis macularia</i> (Spotted sandpiper)	Incubation 21 days max. 8.1 eggs per season usually one brood per season	$C = B = 4$ 1 egg per day	max. 11 eggs per season	>1	F	23, 30
<i>Phalaropus lobatus</i> (Red-necked phalarope)	incubation 17-21 days male care = 33 days usually one brood per season	$C = B = 4$ 1 egg per day	10 days up to 2 clutches per season	>1	F	24, 61
<i>Phalaropus fulicarius</i> (Grey phalarope)	incubation 18-20 days male care = 37 days usually one brood per season	$C = B = 4$	up to 2 per season	>1	F	61, 62
<i>Eudromias morinellus</i> (Dotterel)	incubation 24-28 days male care = 61 days usually one brood per season	$C = B = 3$	5-11 days up to 3 clutches per season	>1	F	61
<i>Jacana spinosa</i> (American jacana)		$C = B = 4$	up to 3 MM per year (≥4 clutches in few weeks')	>1	F	63
<i>Jacana jacana</i> (Wattled jacana)	about 60 days*	$C = B = 4$ 1 egg per day	minimum = 2-4 days up to 6 clutches per season (up to 2 MM)	>1	F	64, 65
<i>Hydrophasianus chirurgus</i> (Pheasant tailed jacana)	about 60 days*	$C = B = 4$	several MM per F at same time	>1	F	66, 67
<i>Metopidius indicus</i> (Bronze winged jacana)	about 60 days*	$C = B = 4^*$	several MM per F at same time	>1	F	68
<i>Actophilornis africana</i> (African jacana)	incubation 24 days male care > 60 days	$C = B = 4$	up to 4 MM per F at same time	>1	F	69, 70
<i>Rostratula benghalensis</i> (Painted snipe)	incubation 15-19 days 1-2 months care (62 days)	$C = B = 4$	up to 4 MM per F at same time	>1	F	69
<i>Turnix sylvaticus</i> (Little button quail)	incubation 12-15 days/care 18-20 days male cycle = 53 days total	$C = B = 3.5$	not known but estimated that F can breed with up to 5 MM per year 10-19 days ($\bar{x} = 13.8$)	>1	F	69, 71
<i>Apogon notatus</i> (Apogonidae: cardinalfish)	8 days care plus 6-15 days = 14-23 days between spawning	$C = B$		>1	F	33
<i>Nerophis ophidion</i> (Syngnathidae: pipefish)	28-37 days	$C = 1.8 \times B$ $M = 204$ eggs per season $F = 396$ eggs per season $C = 1.9 \times B$ $M = 91$ eggs per season $F = 219$ eggs per season		>1 (1.8)	F	25, 31
<i>Syngnathus typhle</i> (Syngnathidae: pipefish)	28-45 days			>1 (1.9)	F	26, 31

Species shown are allocated to two categories on the basis of reports of mating competition: predominant male competitors (a) and predominant female competitors. (b) Although both sexes may compete for particular mating partners, most species could be allocated to the groups without difficulty. Records of reproductive rate ignore the possibility of 'stolen' copulations by males. Abbreviations: M, male; F, female; MM, males; FF, females. Columns show the best available estimates of: (1) the duration of male care of eggs (incubation) and/or young and of the time between successive broods where remating does not follow immediately on independence of the previous brood; (2) average clutch size laid by females (C) and the average brood size cared for by males (B). Where males are known to care for several clutches simultaneously, this is noted; (3) the approximate inter-clutch interval for females or the maximum number of successful breeding partners per season; (4) maximum recorded female reproductive rate (eggs per unit time) divided by the maximum recorded rate of reproduction by males (independent young reared per unit time). For most species, estimates of reproductive rate were only adequate to indicate whether the potential reproductive rate of males exceeded that of females or vice versa; (5) whether males or females are recorded as the primary competitors for mates or breeding territories. * In congeners/other *Jacana* spp.

ratio may be reduced by sex differences in life expectancy⁴¹, which can reflect the costs of increased competition in the potentially faster sex⁷. Variation in the time necessary to find mates may constrain mating competition in some species⁴² (G. Parker, personal communication), while the form of competitive behaviour may be affected by variation in the costs and benefits of particular tactics to the two sexes. Finally, where the potential rate of reproduction is similar in the two sexes, the relative benefits of acquiring qualitatively superior mates^{43,44}, rather than the operational sex ratio, may determine the comparative intensity of mating competition in the two sexes. □

Received 8 October 1990; accepted 5 March 1991.

1. Trivers, R. L. in *Sexual Selection and the Descent of Man* (ed. B. Campbell) 136–179 (Aldine, Chicago, 1972).
2. Trivers, R. L. *Social Evolution* (Cummings, California, 1985).
3. Emlen, S. T. & Oring, L. W. *Science* **197**, 215–223 (1977).
4. Thornhill, R. in *Evolution of Animal Behaviour: Paleontological and Field Approaches* (eds Nitecki, M. H. & Hitchell, J. A.) 113–135 (Oxford University Press, Oxford, 1986).
5. Krebs, J. R. & Davies, N. B. *An Introduction to Animal Ecology* (Blackwell Scientific, Oxford, 1987).
6. Knapton, R. W. *Can. J. Zool.* **62**, 2673–2674 (1984).
7. Clutton-Brock, T. H. *The Evolution of Parental Care* (Princeton University Press, New Jersey, in the press).
8. Gwynne, D. T. *Trends Ecol. Evol.* **6**, 118–121 (1991).
9. Gwynne, D. T. *Behav. Ecol. Sociobiol.* **16**, 355–361 (1985).
10. Gwynne, D. T. & Simmons, L. W. *Nature* **346**, 172–174 (1990).
11. Thornhill, R. & Gwynne, D. T. *Am. Scient.* **74**, 382–389 (1986).
12. Vincent, A. C. J. thesis, Univ. Cambridge (1990).
13. Baylis, J. R. *Nature* **276**, 278 (1978).
14. Baylis, J. R. *Envir. Biol. Fish.* **6**, 223–251 (1981).
15. Breder, C. M. Jr & Rosen, D. E. *Modes of Reproduction in Fishes* (Natural History Press, New York, 1966).
16. Thresher, R. E. *Reproduction in Reef Fishes* (T. F. H. Publications, Neptune City, New Jersey, 1984).
17. Wells, K. D. in *Natural Selection and Social Behavior* (eds Alexander, R. D. & Tinkle, D. W.) 184–197 (Chiron Press, New York, 1981).
18. Nussbaum, R. A. *Misc. Publi. Mus. Zool. Univ. Mich.* **169**, 1–50 (1985).
19. Wilson, E. O. *Insect Societies* (Harvard University Press, Cambridge, Massachusetts, 1971).
20. Thornhill, R. & Alcock, J. *The Evolution of Insect Mating Systems* (Harvard University Press, Cambridge, Massachusetts, 1983).
21. Erckmann, W. J. thesis, Univ. Washington (1981).
22. Handford, P. & Mares, M. A. *Biol. J. Linn. Soc. Lond.* **25**, 77–104 (1985).
23. Oring, L. W., Lank, D. B. & Maxson, S. J. *Auk* **100**, 272–285 (1983).
24. Reynolds, J. D., Colwell, M. A. & Cooke, F. *Behav. Ecol. Sociobiol.* **18**, 303–410 (1986).
25. Rosenqvist, G. *Anim. Behav.* **39**, 1110–1115 (1990).
26. Berglund, A. *Evolution* (in the press).
27. Kluge, A. G. *Misc. Publi. Mus. Zool. Univ. Mich.* **160**, 1–170 (1981).
28. Kynard, B. E. *Behaviour* **67**, 178–207 (1978).
29. Townsend, D. S. *Am. Nat.* **133**, 266–272 (1989).
30. Oring, L. W. & Knudson, M. L. *Living Bird* **11**, 59–73 (1972).
31. Berglund, A., Rosenqvist, G. & Svensson, I. *Am. Nat.* **133**, 506–516 (1989).
32. Berglund, A., Rosenqvist, G. & Svensson, I. *Mar. Ecol. Prog. Ser.* **29**, 209–215 (1986).
33. Kuwamura, T. *Envir. Biol. Fish.* **13**, 17–24 (1985).
34. Bruning, D. F. *Nat. Hist.* **82**, 68–75 (1973).
35. Bruning, D. F. *Living Bird* **13**, 251–294 (1974).
36. Bruning, D. F. thesis, Univ. Colorado (1974).
37. Iwasa, Y., Odendaal, F. J., Murphy, D. D., Ehrlich, P. R. & Launer, A. E. *Theor. Populat. Biol.* **23**, 363–379 (1983).
38. Gregory, P. T. *Can. J. Zool.* **52**, 1063–1069 (1974).
39. Michener, G. R. *Behav. Ecol. Sociobiol.* **14**, 29–38 (1983).
40. Myers, J. P. *Can. J. Zool.* **59**, 1527–1534 (1981).
41. Breitwisch, R. *Curr. Ornithol.* **6**, 1–50 (1989).
42. Sutherland, W. J. *Anim. Behav.* **33**, 1349–1352 (1985).
43. Burley, N. *Proc. natn. Acad. Sci. U.S.A.* **74**, 3476–3479 (1977).
44. Burley, N. *Am. Nat.* **127**, 415–445 (1986).
45. Duellman, W. E. & Trueb, L. *Biology of Amphibians* (McGraw-Hill, New York, 1986).
46. Crespo, J. thesis, Univ. Lisbon (1979).
47. McDiarmid, R. W. in *The Development of Behavior: Comparative and Evolutionary Aspects* (eds Burghardt, C. M. & Bekoff, N.) 127–147 (Garland, New York, 1978).
48. Smith, B. G. *Biol. Bull. Mar. biol. Lab. Woods Hole* **13**, 5–39 (1907).
49. Goto, A. *Copeia* **1987**, 32–40 (1987).
50. Marconato, A. & Bisazza, A. *J. Fish. Biol.* **33**, 905–916 (1988).
51. DeMartini, E. E. *Anim. Behav.* **35**, 1145–1158 (1987).
52. DeMartini, E. E. *Copeia* **1985**, 966–975 (1985).
53. Ochi, H. *Envir. Biol. Fish.* **17**, 117–123 (1986).
54. Gronell, A. M. *Ethology* **81**, 89–122 (1989).
55. Barlow, G. W. *Copeia* **1962**, 346–360 (1962).
56. Barlow, G. W. *Z. Tierpsychol.* **21**, 99–123 (1964).
57. Unger, L. M. *Behav. Ecol. Sociobiol.* **13**, 125–130 (1983).
58. Sargent, R. C. *Behav. Ecol. Sociobiol.* **25**, 379–385 (1989).
59. Gale, W. F. & Deutsch, W. G. *Trans. Am. Fish. Soc.* **224**, 220–229 (1985).
60. Thomson, S. *Anim. Behav.* **34**, 580–589 (1986).
61. Cramp, S. et al. *Handbook of the Birds of Europe, the Middle East and North Africa* Vol. III (Oxford University Press, Oxford, 1983).
62. Ridley, M. W. *Ibis* **122**, 210–226 (1980).
63. Jenni, D. A. & Collier, C. *Auk* **89**, 743–765 (1972).
64. Osborne, D. R. *Wilson Bull.* **94**, 206–208 (1982).
65. Osborne, D. R. & Bourne, G. R. *Condor* **79**, 98–105 (1977).
66. Hoffman, A. *Scop. Zool. Jahrb. Abt. Syst. Oekol. Geogr. Tiere* **78**, 367–403 (1949).
67. Hoffman, A. *Orn. Ber.* **2**, 119–126 (1950).
68. Matthew, D. N. *J. Bombay nat. Hist. Soc.* **61**, 295–301 (1964).
69. Urban, E. K., Fry, C. H. & Keith, S. *The Birds of Africa* Vol. II (Academic, New York, 1986).
70. Vernon, C. J. *Ostrich* **44**, 85 (1973).
71. Wintle, C. C. *Honeyguide* **82**, 27–30 (1975).

ACKNOWLEDGEMENTS. The work arose from discussion of mating competition in pipefish with I. Ahnesjö. For comments, advice, discussion, criticism or access to unpublished material, we thank S. Albon, M. Andersson, A. Balmford, A. Berglund, N. Davies, A. Desrochers, A. Grafen, D. Gwynne, P. Harvey, G. Parker, J. Lazarus, J. Maynard Smith, G. Rosenqvist, and also P. Cassidy for her secretarial help.

Construction of a pattern-generating circuit with neurons of different networks

Pierre Meyrand, John Simmers & Maurice Moulins*

Laboratoire de Neurobiologie et Physiologie Comparées,
Université de Bordeaux I, CNRS, Place du Dr Peyneau,
33120 Arcachon, France

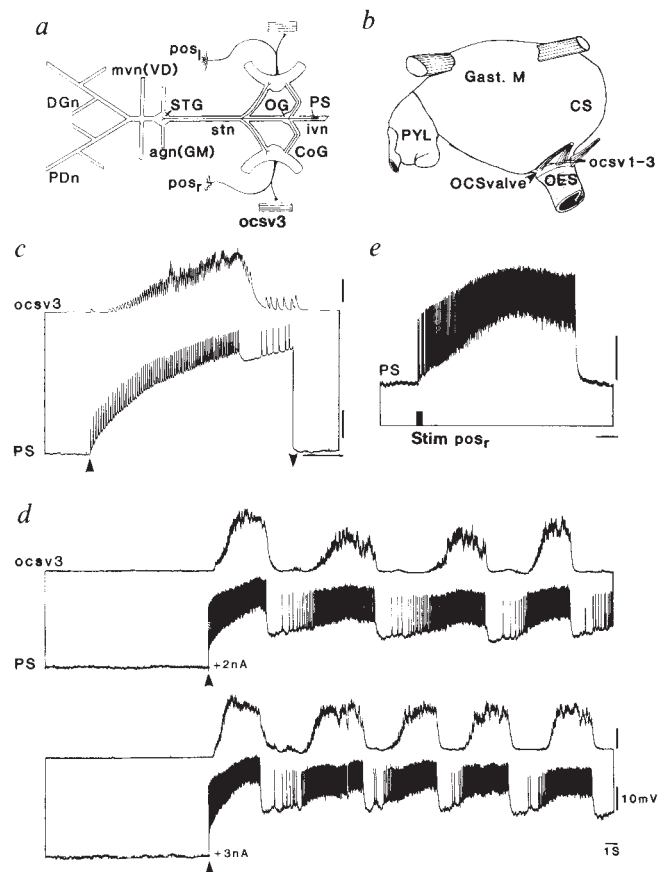
RHYTHMIC motor behaviours are generated within the central nervous system by neuronal circuits called central pattern generators (CPG)¹. Although a CPG can produce several forms of the same behaviour^{2–5} and several circuits may interact to generate different behaviours⁶, it is generally assumed that a given CPG consists of a predefined assemblage of neurons that is functionally distinguishable from other circuits. However, recent studies on the stomatogastric nervous system of crustacea have suggested that CPGs may not be immutable functional entities^{7–10}. We now report that under an identified neuromodulatory stimulus, the CPG that produces swallowing-like behaviour of the foregut in lobsters is constructed *de novo* from neurons belonging to other CPGs. Consequently neurons operating independently as members of different circuits may be reconfigured into a new pattern-generating circuit that operates differently from the original circuits. This not only challenges the concept of the CPG being a discrete functional entity, but also demonstrates that a modulatory input can specify an appropriate CPG from a pool of individual neurons of diverse origins.

We performed our experiments on preparations *in vitro* of the stomatogastric nervous system (STNS) of the lobster *Homarus gammarus*. The STNS consists of four interconnected ganglia (Fig. 1a) that together generate well described motor rhythms of the four regions of the foregut¹¹. These independent foregut rhythms control oesophageal ingestion of food, and its storage in the cardiac sac, trituration by the gastric mill system and filtration through the pylorus on the way to the midgut (Fig. 1b). We describe here a new distinct motor activity of the STNS which transfers food between these different foregut compartments. We show that this swallowing-like behaviour arises first from the rhythmic opening of a valve situated between the oesophagus and the cardiac sac (OCS valve) (Fig. 1b), and second from a massive reorganization of all other foregut rhythms (Figs 2 and 3).

In STNs preparations with the anterior part of the foregut left attached, the three dilator muscles (ocsv1–3; Fig. 1b) of the OCS valve are generally inactive and the latter remains closed. We have found that opening of the valve is driven from the commissural ganglion (Fig. 1a) by OCS dilator motoneurons, which in turn are controlled by two equivalent interneurons arising in the inferior ventricular nerve (Fig. 1a). These cells have been previously identified and named 'pyloric suppressors'¹² (PS). Intracellular depolarization of either PS neuron to evoke firing strongly activates ocsv dilator muscles (Fig. 1c). Although PS is generally silent *in vitro*, we believe the neuron has endogenous bursting properties that drive rhythmic dilation

* To whom correspondence should be addressed.

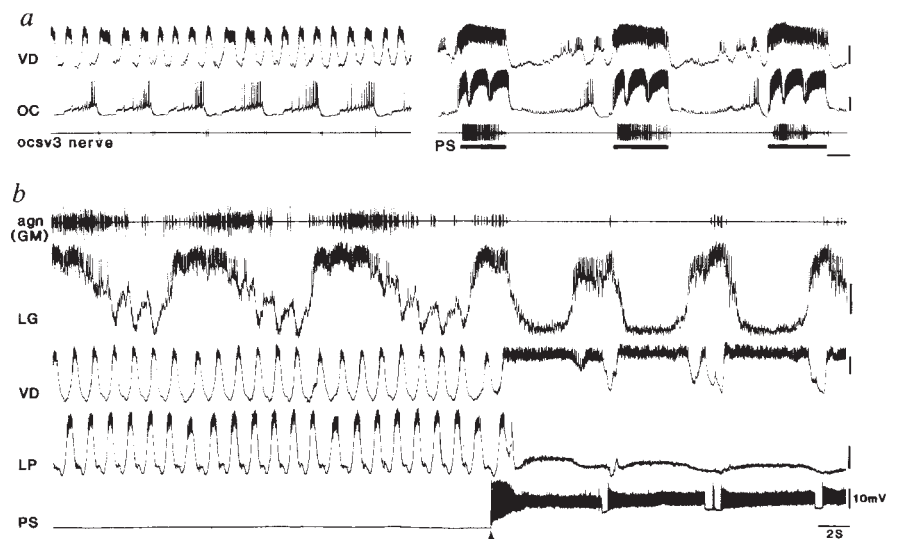
FIG. 1 Neuron pyloric suppressor (PS) drives rhythmic dilation of the oesophageal/cardiac sac valve (OCS) through its endogenous bursting properties. *a*, The isolated stomatogastric nervous system of *Homarus gammarus* showing the soma position and axonal projections of interneuron PS. *b*, Lateral view of the foregut showing the position of the OCS valve and the four functionally separate regions of the foregut. In *a* and *b*, anterior is to the right. *c*, Simultaneous intracellular recordings from the soma of PS and a fibre of *oscv*₃, a dilator muscle of the OCS valve. Discharge of neuron PS, here evoked by intrasomatic injection of depolarizing current (arrowheads), excites the dilator motoneuron to the OCS valve and causes synaptic depolarization of muscle *oscv*₃. *d*, When continuously depolarized (current level 2 nA, upper panel; 3 nA, lower panel), an otherwise silent PS fires in repetitive spike bursts and rhythmically drives the dilator muscle of the OCS valve. The higher burst frequency in the lower panel indicates a voltage-dependent oscillatory mechanism intrinsic to PS itself. *e*, Sensory activation of neuron PS. A brief electrical stimulation (30 Hz for 250 ms) of the nerve branch to the right posterior oesophageal sensor (see *a*), a chemosensor in the internal wall of the OCS valve¹⁴, triggers an active spike burst in PS that long outlasts the initial stimulus. Abbreviations: agn, anterior gastric nerve; CoG, commissural ganglion; CS, cardiac sac; DGn, nerve containing the axon of DG motoneuron; Gast. M, gastric mill; GM, gastric mill motoneuron; ivn, inferior ventricular nerve; mvn, median ventricular nerve; OCS valve, oesophageal-cardiac sac valve; *oscv*₁₋₃, dilator muscles 1-3 of the OCS valve; OG, oesophageal ganglion; OES, oesophagus; *pos*_l and *pos*_r, left and right posterior oesophageal sensors; PDn, nerve containing the axon of PD motoneuron; PS, pyloric suppressor neuron; PYL, pyloric chamber; STG, stomatogastric ganglion; stn, stomatogastric nerve; VD, ventral dilator motoneuron. Dissections and electrophysiological recording and stimulation were made as previously described for this system¹¹.



of the OCS valve. First the cell always oscillates and generates repetitive bursts during occasional spontaneous bouts of activity, or in response to sustained depolarizing current (Fig. 1*d*). In either case, the *oscv* dilator muscles are active in time with the bursts of PS. Second, bursting of PS is voltage-dependent, displaying a threshold for activation and an increase in frequency with increasing levels of tonic depolarization (Fig. 1*d*). Third, regenerative bursts in PS can be triggered synaptically by brief electrical stimulation of bilateral sensory nerves (Fig. 1*e*) from the posterior oesophageal sensors (Fig. 1*a*), also suggesting an input pathway appropriate for the induction of swallowing behaviour *in vivo*.

Neuron PS also makes synaptic connections with neurons of other STNS circuits; pyloric cells were previously shown to receive inputs from PS¹³, and we have found similar direct connections onto gastric and oesophageal neurons (P.M., J.S. and M.M., manuscript in preparation). We therefore assessed the contribution of these individual circuits to swallowing behaviour. In contrast to PS neurons, all three circuits are continuously active *in vitro*. Figure 2*a* (left panel) shows spontaneous oesophageal (monitored by neuron OC) and pyloric (neuron VD) rhythms in the absence of PS activity. Figure 2*b* (left) shows a typical gastric rhythm (neurons GM, LG) recorded simultaneously with neurons VD and LP of the pyloric circuit.

FIG. 2 Discharge of neuron PS restructures independent stomatogastric rhythms into a single novel motor pattern. *a*, left panel, Spontaneous pyloric and oesophageal circuit activity recorded simultaneously from neurons VD and OC, respectively. The oesophageal rhythm typically has a longer cycle period (4-6 s) than the pyloric rhythm (1-2 s). The extracellularly recorded dilator motoneuron to the OCS valve (*oscv*₃) was weakly active in oesophageal time. *a*, right panel, Same preparation during spontaneous bursting in PS (cell not penetrated in this experiment; bars indicate timing of PS bursts). The OCS valve dilator motoneuron is now strongly activated by PS (as in Fig. 1*d*), and both VD and OC of the previously separate pyloric and oesophageal patterns become coordinated to this new rhythm. *b*, PS acts differently on individual stomatogastric neurons. Left, independent pyloric (VD, LP neurons) and gastric (LG, GM in agn) circuit activities in the absence of PS firing. The cycle period of the gastric rhythm is typically 10-15 s. When PS is strongly depolarized (arrowhead) and fires in long repetitive bursts, gastric LG and pyloric VD immediately combine to produce a new single rhythm in time with the interneuron, while activity in gastric GM and pyloric LP is immediately decreased or completely suppressed. Abbreviations: LG, lateral gastric motoneuron; LP, lateral pyloric motoneuron; OC, oesophageal constrictor motoneuron; other abbreviations as in legend to Fig. 1.



ations: LG, lateral gastric motoneuron; LP, lateral pyloric motoneuron; OC, oesophageal constrictor motoneuron; other abbreviations as in legend to Fig. 1.

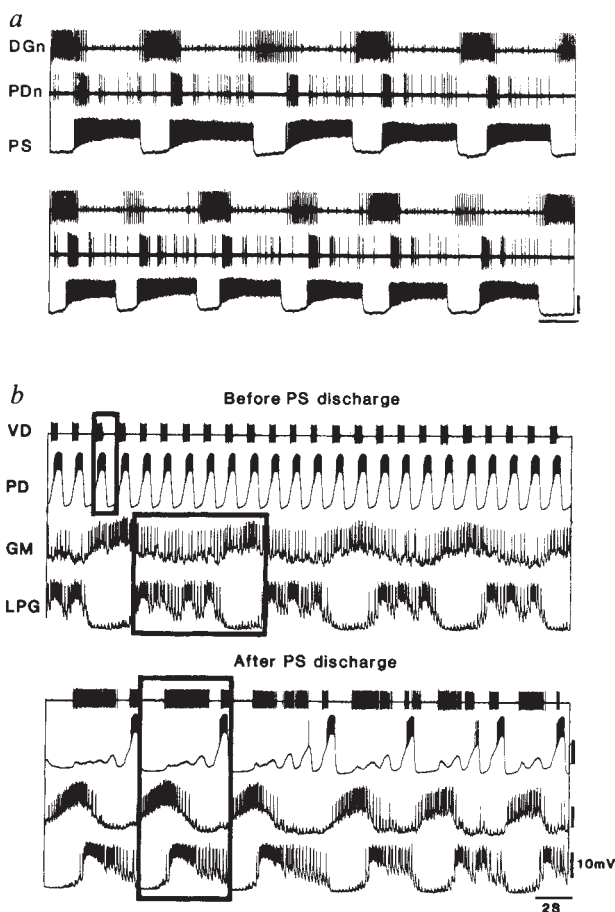


FIG. 3 Neuron PS imposes its own rhythm on the swallowing motor pattern and produces a long-lasting reconfiguration of stomatogastric circuits. *a*, PS-driven swallowing activity in the gastric and pyloric circuits monitored extracellularly from DG (DGn) and PD (PDn) motoneurons, respectively. The two panels show membrane potential oscillation and bursting in the same PS neuron in response to two different levels (3 nA, upper panel; 4 nA, lower panel) of tonic depolarizing current. The activity of both networks follow the voltage-dependent changes in rhythm frequency of PS. *b*, Simultaneous recordings from neurons of the pyloric (VD, PD) and gastric (GM, LPG) networks before and immediately after a 10 s burst (mean spike frequency 25 Hz) in PS. Upper panel, separate pyloric and gastric rhythms (rectangles) before PS firing. Lower panel, after PS discharge, the same neurons are coordinated in a single rhythm (rectangle) that is different from either of the two original patterns. Abbreviations: LPG, lateral posterior gastric motoneuron; PD, pyloric dilator motoneuron; other abbreviations as in legend to Fig. 1.

When the interneuron becomes active, however, dramatic alterations occur in these otherwise independent and very different motor patterns. First, the ongoing activity of all three circuits is reconfigured into a single novel pattern. The rhythms of neurons OC and VD become coordinated to spontaneous burst-

ing in PS (here monitored by recording from the dilator nerve to the OCS valve; Fig. 2*a*, right panel); neuron VD now fires in single long bursts in time with PS while oesophageal constrictor neuron OC discharges rhythmically in high frequency bursts. Both cells remain inactive throughout most of the interneuron's interburst interval and the new rhythm bears no resemblance to either of the original oesophageal and pyloric rhythms. Similarly in Fig. 2*b* (right), current-induced bursting in PS produces a new combined pattern consisting of pyloric neuron VD firing in time with the interneuron, but alternately with the LG neuron, an integral member of the gastric circuit.

Second, neuron PS acts differently on individual elements of each circuit. Although certain neurons are strongly activated in phase (for example neurons VD and OC in Fig. 2*a*, right panel) or in antiphase (such as neuron LG and VD in Fig. 2*b*, right) with the interneuron, other cells are strongly inactivated (neurons GM and LP in Fig. 2*b*) and take no part in the swallowing motor pattern. Thus the interneuron can selectively discard integral members of different STNS circuits and assemble the remaining subpopulation of neurons into a new functional network.

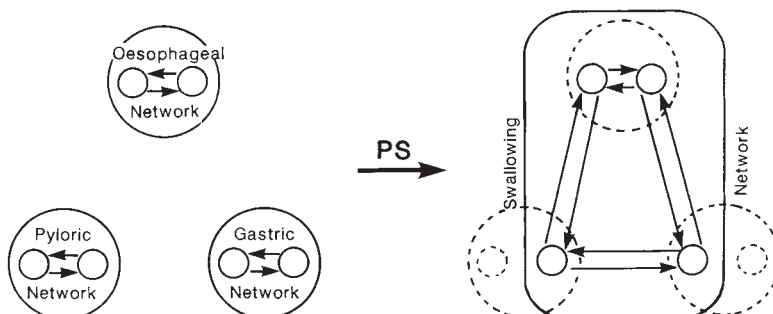
Third, the PS neuron imposes its own rhythm on this swallowing network, thereby timing its activity to dilation of the OCS valve. This is evident in Fig. 3*a*, where PS is induced to burst at two different cycle frequencies by different levels of injected depolarizing current. In each case (Fig. 3*a*, upper and lower panels) activity in gastric neuron DG and pyloric neuron PD, like the dilator motoneuron of the OCS valve (Fig. 1*d*), follow closely the inherent rhythm of PS.

Fourth, neuron PS exerts long-lasting influences on STNS activity (Fig. 3*b*). Although original oesophageal rhythmicity returns soon after PS discharge (not shown), neurons ordinarily belonging to the gastric (GM, LPG) and pyloric (VD, PD) networks (Fig. 3*b*, upper panel) remain coordinated in a single pattern (Fig. 3*b* lower panel) for further tens of seconds before they again recover their separate gastric and pyloric network identities.

This reconfiguration of entire foregut motor activity seems to be behaviourally relevant. Our data imply that when the OCS valve dilates cyclically *in vivo* (during an episode of PS bursting; Fig. 1*d*): (1) a strong increase in oesophageal rhythmicity during each opening of the valve (Fig. 2*a*, right panel) serves to push food into the stomach; and (2) a unified cycling pattern, also synchronized to valve openings, occurring in the more posterior regions of the foregut (Fig. 2*b* and 3*a*), facilitates the rearward transit of the food. This cooperative swallowing pattern persists for some time after closure of the OCS valve (PS neuron again silent; Fig. 3*b* lower panel), before the flow of ingested food terminates and each foregut region resumes its separate motor function.

The effects of PS on individual pyloric, oesophageal and gastric neurons suggest several mechanisms by which the cell reconfigures the various STNS networks. First, PS modulates intrinsic membrane characteristics in several target neurons (such as causing the long-term suppression of oscillatory properties in pyloric cell LP (ref. 13; Fig. 2*b*) and so considerably

FIG. 4 Diagrammatic representation of PS-induced construction of the swallowing neural network from individual circuits of the stomatogastric nervous system. Left; with no PS activity, the oesophageal, pyloric and gastric pattern generators operate independently at significantly different cycle periods and control their respective regions of the foregut. Right; rhythmic discharge of PS drives opening of the OCS valve and, by breaking down the pyloric and gastric circuits and superimposing its own rhythm, the interneuron builds a single new network appropriate for swallowing behaviour.



disrupts the ability of the parent network to produce its inherent rhythm. Second, PS also acts directly on many oesophageal, gastric and pyloric neurons by way of conventional excitatory and inhibitory synapses¹³, serving to override remaining intranetwork interactions and couple these elements to the interneuron's own rhythm (Figs 2 and 3a). It is also probable that the neuron PS modulates and potentiates normally weak synaptic interactions⁹, such as those between pyloric and gastric neurons¹¹, thereby explaining the coordination of these networks that persists after the interneuron falls silent (Fig. 3b).

Figure 4 summarizes these results which show that expression of a behaviour can be obtained by a complete restructuring of several networks that otherwise produce different behaviours. Although different patterns of motor output may arise from

changes within individual neural networks^{3,4}, only recently has such functional reconfiguration been found to extend to apparently separate networks composed of different neurons. In lobster and crab STNS, individual neurons can switch between different rhythm-generating networks^{7,8,10} and two independent networks can merge into a single functional unit⁹. In our study, rather than just a switch in neuronal 'allegiance' to two ongoing rhythms, or 'fusion' of whole active networks to produce a combined rhythm, we find the selective dismantling of preexisting networks and the construction of a completely new motor circuit. This indicates that a pattern-generating network may exist only in a particular behavioural situation dictated by neuromodulatory influences. □

Received 31 December 1990; accepted 4 March 1991.

1. Delcomyn, F. *Science* **210**, 492-498 (1980).
2. Nagy, F. & Moulins, M. in *The Crustacean Stomatogastric System* (eds Selverston, A. I. & Moulins, M.) 205-262 (Springer, Berlin, 1987).
3. Marder, E. *Nature* **335**, 296-297 (1988).
4. Harris-Warrick, R. M. in *Neuronal Control of Rhythmic Movements in Vertebrates* (eds Cohen, A. V., Rossignol, S. & Grillner, S.) 285-331 (Wiley, New York, 1988).
5. Stein, P. S. G., Mortin, L. I. & Robertson, G. A. in *Neurobiology of Vertebrate Locomotion* (eds Grillner, S. et al.) 201-216 (Macmillan, London, 1988).

6. Grillner, S. *Science* **228**, 143-149 (1985).
7. Meyrand, P., Weimann, J. M. & Marder, E. *Soc. Neurosci. Abstr.* **14**, (1988).
8. Hooper, S. L. & Moulins, M. *Science* **244**, 1587-1589 (1989).
9. Dickinson, P. S., Mecsas, C. & Marder, E. *Nature* **344**, 155-158 (1990).
10. Weimann, J. M., Meyrand, P. & Marder, E. *J. Neurophysiol.* **65**, 111-122 (1991).
11. Selverston, A. I. & Moulins, M. (eds) *The Crustacean Stomatogastric System* (Springer, Berlin, 1987).
12. Cazalets, J. R., Nagy, F. & Moulins, M. *J. Neurosci.* **10**, 448-457 (1990).
13. Cazalets, J. R., Nagy, F. & Moulins, M. *J. Neurosci.* **10**, 458-468 (1990).
14. Robertson, R. M. & Laverack, M. S. *Proc. R. Soc. B* **206**, 235-263 (1979).

Mutant α subunits of G_{12} inhibit cyclic AMP accumulation

Yung H. Wong, Alex Federman, Ann M. Pace, Ian Zachary*, Tony Evans†, Jacques Pouyssegur‡ & Henry R. Bourne§

Departments of Pharmacology and Medicine and the Cardiovascular Research Institute, University of California, San Francisco, California 94143, USA

† Genentech Inc., South San Francisco, California 94080, USA

‡ Centre de Biochimie du CNRS, Faculté des Sciences, Université de Nice, 06034 Nice, France

ONE or more of three G_i proteins, G_{i1-3} , mediates hormonal inhibition of adenylyl cyclase¹⁻³. Whether this inhibition is mediated by the α or by the $\beta\gamma$ subunits of G_i proteins is unclear^{1,2}. Mutations inhibiting the intrinsic GTPase activity of another G protein, the stimulatory regulator of adenylyl cyclase (G_s), constitutively activate it by replacing either of two conserved amino acids in its α subunit (α_s)⁴⁻⁷. These mutations create the *gsp* oncogene which is found in human pituitary and thyroid tumours^{5,8}. In a second group of human endocrine tumours, somatic mutations in the α subunit of G_{12} replace a residue cognate to one of those affected by *gsp* mutations⁸. This implies that the mutations convert the α_{12} gene into a dominantly acting oncogene, called *gip2* (ref. 8), and that the mutant α_{12} subunits are constitutively active. We have therefore assessed cyclic AMP accumulation in cultured cells which stably or transiently express exogenous wild-type α_{12} complementary DNA or either of two mutant α_{12} cDNAs. The results show that putatively oncogenic mutations in α_{12} constitutively activate the protein's ability to inhibit cAMP accumulation.

We mutated mouse α_{12} at codons cognate to either of the two α_s codons affected by *gsp* mutations^{5,8}, producing cDNAs that encode leucine instead of glutamine at position 205 (α_{12} -Q205L) or cysteine instead of arginine at position 179 (α_{12} -R179C). Wild-type (WT) and mutant G protein α chains were stably expressed in NIH3T3 cells using retroviral infection with a

vector containing a dominant selectable marker, *neo*, which confers resistance to G418. Infected G418-resistant pools were tested for cAMP accumulation in response to prostaglandin E_1 (PGE_1), which stimulates adenylyl cyclase by way of receptors coupled to G_s , or to forskolin, which is thought to stimulate adenylyl cyclase directly⁹. In control cells (infected with vector alone or with vectors expressing α_s -WT or α_{12} -WT), PGE_1 and forskolin elevated cellular cAMP by about 4- and 20-fold, respectively (Table 1). As in other cells^{10,11}, lysophosphatidic acid (LPA) inhibited PGE_1 - and forskolin-stimulated cAMP accumulation by 30-50% in all the control populations of NIH-3T3 cells (Table 1). Treatment with pertussis toxin (PTX; 100 ng ml⁻¹ for 20 h) abolished the inhibitory effect of LPA (result not shown), suggesting that it is mediated by a G_i -like protein.

Mutationally activated α_s (α_s -Q227L) constitutively increased basal and stimulated cAMP accumulation (Table 1; refs 5 and 12). Expression of the two mutant α_{12} subunits also produced a dominant, constitutive effect on cAMP accumulation, albeit in the opposite direction. Expression of either α_{12} -R179C or

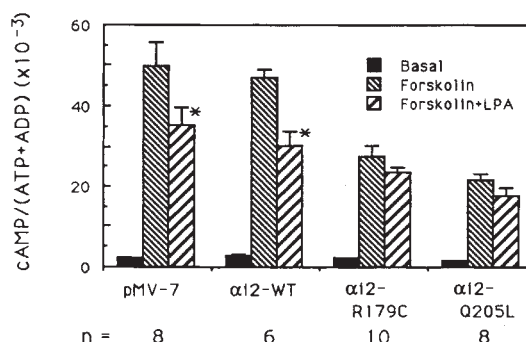


FIG. 1 cAMP accumulation in multiple individual clones derived from G418-selected pools of NIH3T3 cells infected with vector (pMV-7), or vector containing α_{12} -WT, α_{12} -Q205L, or α_{12} -R179C. Clonal lines were generated by limiting dilution and cAMP accumulation was measured as described in the legend to Table 1. Forskolin was used at 50 μ M, LPA at 100 μ M. The results represent the mean $\pm$ s.e.m. of the responses obtained from the number of clones indicated in parentheses. *LPA significantly reduced the forskolin-stimulated activity; paired *t*-test, $P < 0.05$.

* Present address: Imperial Cancer Research Fund, Lincoln's Inn Fields, London WC2A 3PX, UK.
§ To whom correspondence should be addressed.

TABLE 1 cAMP accumulation in pools of infected NIH-3T3 cells

	cAMP/(ATP + ADP) ($\times 10^{-3}$)				
	Basal	PGE ₁ *	PGE ₁ + LPA†	Forskolin*	Forskolin + LPA†
NIH-3T3	1.8 ± 0.1	6.6 ± 0.2	4.4 ± 0.35	47.6 ± 2.0	30.8 ± 1.3
pMV-7	2.0 ± 0.2	7.9 ± 0.4	4.5 ± 0.4	47.4 ± 5.5	31.0 ± 2.2
α_s -WT	2.1 ± 0.1	7.2 ± 0.4	4.1 ± 0.3	45.5 ± 2.7	29.4 ± 3.9
α_s -Q227L	4.2 ± 0.7‡	10.0 ± 1.4	8.1 ± 0.9§	135 ± 15.3‡	91.1 ± 12.1
α_{12} -WT	1.7 ± 0.1	6.3 ± 0.1	4.5 ± 0.3	46.0 ± 0.4	32.1 ± 2.0
α_{12} -Q205L	1.5 ± 0.1	3.1 ± 0.4‡	2.6 ± 0.2§	21.3 ± 1.8‡	17.6 ± 1.4§
α_{12} -R179C	1.7 ± 0.1	4.8 ± 0.2‡	3.2 ± 0.1§	31.7 ± 3.1‡	24.4 ± 4.6§

[³H]cAMP in [³H]adenine-labelled NIH3T3 fibroblasts infected with either vector alone or vectors containing the indicated cDNA was estimated by determining the ratios of cAMP to total ATP and ADP pools. PGE₁ and forskolin were present at 50 μ M, LPA at 100 μ M. Values represent the mean $\pm$ s.e.m. of 3–5 independent experiments in triplicate determinations. Mouse α_{12} cDNA²⁶ was subjected to site-directed mutagenesis, using the Bio-Rad Muta-Gene kit. Wild-type and mutant cDNAs were subcloned into the retroviral vector pMV-7, processed through packaging cell lines and used to infect NIH-3T3 cells, as described²⁷. Cells were cultured in Dulbecco's modified Eagle's medium (DMEM) containing 10% calf serum and infectants were selected in the same medium supplemented with 500 μ g per ml G418. For measuring intracellular cAMP, confluent cells in 24-well plates prelabelled with [³H]adenine (2 μ Ci ml⁻¹ for 20–24 h) were washed once with HEPES-buffered DMEM and incubated (37 °C for 30 min) in the same medium containing 1 mM 1-methyl-3-isobutylxanthine with or without the indicated drugs. Reactions were terminated by aspiration and the immediate addition of 5% trichloroacetic acid (1 ml per well). Acid-soluble nucleotides were separated on ion-exchange columns as described²⁸.

* Stimulatory responses to PGE₁ or forskolin were significantly higher than the corresponding basal values; paired *t*-test, *P* < 0.05.

† Unless otherwise stated, responses to PGE₁ or forskolin were significantly inhibited in the presence of LPA; paired *t*-test, *P* < 0.05.

‡ Significantly different from uninfected or vector infected cells; *t*-test, *P* < 0.05.

§ Inhibitory responses produced by LPA were statistically insignificant; *t*-test.

α_{12} -Q205L decreased both PGE₁- and forskolin-stimulated cAMP accumulation in pools of G418-resistant cells (Table 1) and in multiple individual clones generated from pools of infected G418 resistant cells (Fig. 1); α_{12} -WT did not affect cAMP accumulation.

The inhibitory effects of α_{12} -R179C and α_{12} -Q205L lowered agonist-stimulated cAMP accumulation to levels at or below those produced by the combination of stimulatory agonist plus LPA in control cells, suggesting that both mutations constitu-

tively activate α_{12} . Inhibition by mutationally activated α_{12} seems to mimic G_i-mediated inhibition of cAMP accumulation by LPA; the two inhibitory effects are not additive (Table 1; Fig. 1).

To extend this investigation, we examined hormonal regulation of cAMP accumulation in a human embryo kidney cell line, 293, cotransfected with the cDNA to be tested and with a vector expressing the cDNA for the rat luteinizing hormone receptor (LHR)¹³, which stimulates adenylyl cyclase through G_s. Cotransfection with cDNA encoding the porcine α_2 adrenoreceptor¹⁴ allowed an α_2 adrenoreceptor agonist, UK-14304, to reduce the cAMP accumulation stimulated by an LHR agonist, human chorionic gonadotropin (hCG), by 60% (Fig. 2a). The effect of UK-14304, like that of hCG, required expression of the appropriate receptor. Therefore, the stimulatory and inhibitory receptors are expressed in the same subpopulation of 293 cells.

In this model system, transient expression of either α_{12} -R179C or α_{12} -Q205L inhibited hCG-stimulated cAMP accumulation. These inhibitory effects increased with increasing mutant α_{12} DNA (Fig. 2b) and were similar to those mediated by the α_2 adrenoreceptor. Inhibition by mutant α_{12} was not additive to that mediated by the α_2 adrenoreceptor (result not shown). Although PTX prevented inhibition of cAMP accumulation by UK-14304 and LPA in 293 and NIH3T3 cells, respectively, the toxin did not affect inhibition of cAMP accumulation by mutant α_{12} in either 293 (Fig. 2b) or NIH3T3 cells (results not shown). The mutant α_{12} chains were nonetheless good substrates for PTX-catalysed ADP-ribosylation (Fig. 3), consistent with observations^{2,15,16} that covalent modification by PTX preferentially impairs the interactions of α subunits with receptors, rather than effectors.

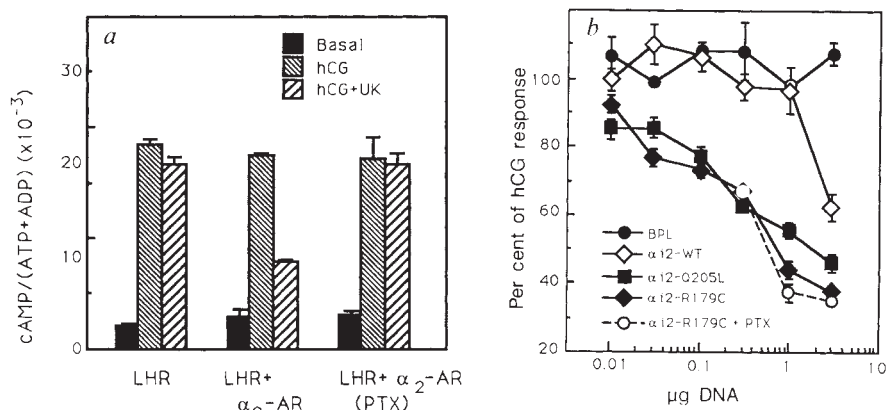
Effects of two control DNAs, encoding bovine prolactin or α_{12} -WT, indicated that the inhibitory effects of mutant α_{12} were specific in 293 cells (Fig. 2b). α_{12} -WT inhibited cAMP accumulation only at the highest DNA concentration tested.

Our observations in NIH3T3 and 293 cells indicate that α_{12} mutations designed to mimic GTPase-inhibiting mutations of α_s constitutively activate the protein, so that cAMP accumulation is inhibited. These results reinforce the view that G protein α chains use similar three-dimensional structures and the same molecular mechanism to hydrolyse GTP. Specifically, this probably involves the side chains of two conserved amino acids in all α chains: A glutamine equivalent to positions 205 of α_{12} or 227 of α_s and an arginine equivalent to positions 179 and 201 of the same proteins.

Others^{1,17} have proposed that G_i mediates hormonal inhibi-

FIG. 2 cAMP accumulation in 293 cells. a, Inhibition of the stimulatory effect of hCG (5 ng ml⁻¹) by an α_2 adrenoreceptor agonist, UK-14304 (10 nM), in cells transfected with DNA encoding LHR (0.5 μ g DNA per 10⁶ cells), or LHR plus α_2 adrenoreceptor (α_2 -AR; 0.05 μ g DNA per 10⁶ cells). b, Inhibition of hCG-stimulated cAMP accumulation by wild-type and mutant α_{12} in cells cotransfected with DNA containing LHR (0.5 μ g DNA per 10⁶ cells) and the indicated amounts of DNA containing α_{12} -WT, α_{12} -Q205L, α_{12} -R179C, or bovine prolactin (BPL). Results are expressed as a per cent of the hCG-stimulated (5 ng ml⁻¹) activity as compared with that measured in cells transfected with LHR alone. The data represent triplicate determinations in one experiment; two additional experiments gave similar results.

METHODS. Cells were maintained in DMEM containing 10% fetal calf serum. Transient expression of LHR and α_2 adrenoreceptor cDNAs in the vectors pCIS and pCMV4, respectively, has been described^{13,14}. Wild-type and mutant α_{12} cDNAs were subcloned into the eukaryotic expression vector pcDNA1 (Invitrogen) and transfected into 293 cells (10⁶ cells on 60-mm culture dish) by calcium phosphate precipitation. Sixteen hours after transfection, cells were rinsed once with Waymouth's medium



containing 1 mg per ml BSA and re-seeded in 12-well plates. Cells were then labelled with [³H]adenine for 24 h, in the absence or presence of PTX (100 ng ml⁻¹), and cAMP accumulation was assayed (48 h after transfection) in the presence of 1 mM 1-methyl-3-isobutylxanthine and the indicated drugs as described in the legend to Table 1.

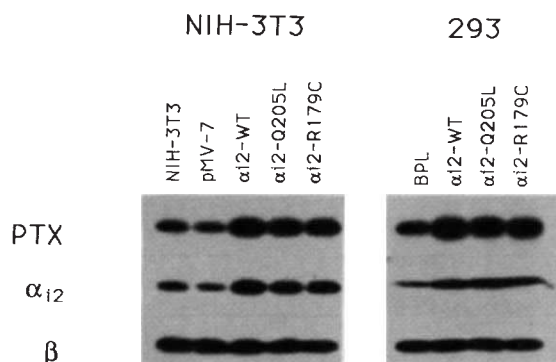


FIG. 3 Detection of α_{i2} and β subunits in cells expressing wild-type and mutant α_{i2} . Membrane preparations were subjected either to immunoblot analysis using antisera specific to α_{i2} or β subunits, or to [32 P]ADP-ribosylation catalysed by PTX. Membranes were prepared from representative clones of NIH3T3 cells isolated from pooled cell populations infected with the indicated DNA (left) or from transfected 293 cells (right). The 293 cells were transfected with 0.5 μ g LHR DNA and 0.3 μ g α_{i2} or control DNA (encoding bovine prolactin) per 10^6 cells.

METHODS. Cells were frozen and thawed once and then homogenized by 10 passages through a 27-gauge needle. Nuclei were removed by low-speed centrifugation and a membrane fraction isolated by centrifugation for 30 min in a microfuge. Membrane proteins (100 μ g) were resolved on 12.5% SDS-polyacrylamide gels and transferred to Immobilon-P membranes (Millipore). Detection of the antibody-antigen complex was by 125 I-labelled protein A. Rabbit polyclonal antibodies were raised against the α_{i2} peptide sequence CAEEQGMLPEDLSG (residues 112–126) and the specificity confirmed by immunoblotting of purified α_i subunits isolated from bovine brain or human platelets (data not shown). The antisera were affinity-purified before use. The specificity of antiserum Beta-8 for the detection of β subunits was as previously described²⁹. PTX-catalysed [32 P]ADP-ribosylation reactions using 50 μ g membrane protein were performed exactly as described³⁰. Films were exposed for 2–3 days for immunoblots, and 2–4 h for ADP-ribosylation.

tion of adenylyl cyclase by a mechanism in which $\beta\gamma$ first dissociates from α_i and then binds to and inactivates α_s . This is consistent with the ability of $\beta\gamma$ to inhibit α_s (ref. 18) and adenylyl cyclase activity *in vitro*. But other evidence^{16,19,20} suggests that α_i can directly inhibit adenylyl cyclase, without help from $\beta\gamma$. Moreover, titration of α_s by $\beta\gamma$ cannot be the sole mechanism by which G_i mediates hormonal inhibition of adenylyl cyclase, because GTP analogues and somatostatin inhibit adenylyl cyclase in mutant S49 lymphoma cells totally lacking α_s (refs 21–25).

Although the $\beta\gamma$ subunit of G_i may mediate hormonal inhibition of adenylyl cyclase in certain circumstances, inhibition of cAMP accumulation in our experiments is difficult to explain as an effect of $\beta\gamma$ and is much more likely to be mediated by α_{i2} . The amount of immunoreactive β does not detectably increase in either NIH3T3 or 293 cells expressing exogenous α_{i2} , whereas the 'extra' mutant (or wild-type) α_{i2} can easily be detected by immunoblots and PTX-catalysed ADP-ribosylation (Fig. 3).

Our experiments show that mutations in the glutamine-205 and arginine-179 codons of α_{i2} produce a protein that constitutively inhibits cAMP accumulation in cultured cells. Although the mechanism of this inhibition has not been determined, it probably involves constitutive activation of the α_{i2} chain. These conclusions support the proposal⁸ that α_{i2} mutations create a dominantly acting *gip2* oncogene, although they do not necessarily imply that decreased cAMP accumulation mediates the oncogenic effect of *gip2*. □

- Landis, C. A. *et al.* *Nature* **340**, 692–696 (1989).
- Graziano, M. P. & Gilman, A. G. *J. biol. Chem.* **264**, 15475–15482 (1989).
- Freissmuth, M. & Gilman, A. G. *J. biol. Chem.* **264**, 21907–21914 (1989).
- Lyons, J. *et al.* *Science* **249**, 655–659 (1990).
- Seamon, K. & Daly, J. W. *J. biol. Chem.* **257**, 9799–9801 (1981).
- Murayama, T. & Ui, M. *J. biol. Chem.* **262**, 5522–5529 (1987).
- van Corven, E. J., Groenink, A., Jalink, K., Eichholtz, T. & Moolenaar, W. H. *Cell* **59**, 45–54 (1989).
- Zachary, I., Masters, S. B. & Bourne, H. R. *Biochem. biophys. Res. Commun.* **168**, 1184–1193 (1990).
- McFarland, K. C. *et al.* *Science* **245**, 494–499 (1989).
- Guyer, C. A. *et al.* *J. biol. Chem.* **265**, 17307–17317 (1990).
- Price, S. R., Barber, A. & Moss, J. in *ADP-ribosylating Toxins and G Proteins. Insights into Signal Transduction* (eds Moss, J. & Vaughan, M.) 397–424 (Am. Soc. Microbiology, Washington, DC, 1990).
- Katada, T., Oinuma, M. & Ui, M. *J. biol. Chem.* **261**, 5215–5221 (1986).
- Gilman, A. G. *Rev. Biochem.* **56**, 615–649 (1987).
- Katada, T., Northup, J. K., Bokoch, G. M., Ui, M. & Gilman, A. G. *J. biol. Chem.* **259**, 3578–3585 (1984).
- Hildebrandt, J. D., Codina, J. & Birnbaumer, L. *J. biol. Chem.* **259**, 13178–13185 (1984).
- Kobayashi, I. *et al.* *Eur. J. Biochem.* **191**, 499–506 (1990).
- Hildebrandt, J. D., Hanoune, J. & Birnbaumer, L. *J. biol. Chem.* **257**, 14723–14725 (1982).
- Hildebrandt, J. D. *et al.* *Nature* **302**, 706–709 (1983).
- Jakobs, K. H., Aktories, K. & Schultz, G. *Nature* **303**, 177–178 (1983).
- Jakobs, K. H. & Schultz, G. *Proc. natn. Acad. Sci. U.S.A.* **80**, 3899–3902 (1983).
- Katada, T., Bokoch, G. M., Smigel, M. D., Ui, M. & Gilman, A. G. *J. biol. Chem.* **259**, 3586–3595 (1984).
- Sullivan, K. A., *et al.* *Proc. natn. Acad. Sci. U.S.A.* **83**, 6687–6691 (1986).
- Sullivan, K. A., *et al.* *Nature* **330**, 758–760 (1987).
- Salomon, Y., Londos, C. & Rodbell, M. *Analyt. Biochem.* **58**, 541–548 (1974).
- Evans, T., Fawzi, A., Fraser, E. D., Brown, M. L. & Northup, J. K. *J. biol. Chem.* **262**, 176–181 (1987).
- Chang, F.-H. & Bourne, H. R. *Endocrinology* **121**, 1711–1715 (1987).

ACKNOWLEDGEMENTS. We thank D. Segaloff for 293 cells and for the LHR cDNA in the pCIS vector, D. Horstman for the α_2 adrenoceptor cDNA in the pCMV4 vector and D. Julius for NIH-3T3 cells. This work was supported in part by grants from the NIH and the March of Dimes.

Early aspects of *Caenorhabditis elegans* sex determination and dosage compensation are regulated by a zinc-finger protein

Michael L. Nonet & Barbara J. Meyer*

Department of Molecular and Cell Biology, University of California, Berkeley, California 94720, USA

THE *sd-1* gene acts at an early step in the regulatory hierarchy that controls the choice of sexual fate in *Caenorhabditis elegans*. It functions at a point before the control of sex determination and X-chromosome dosage compensation diverge. Here we report that *sd-1* encodes a protein of 1,203 amino acids containing seven zinc fingers. This protein motif in combination with other genetic and molecular information suggests that *sd-1* is likely to function as an embryonic transcription factor regulating downstream genes involved specifically in the sex determination and dosage compensation pathways, or regulating other genes involved in the coordinate control of both processes. These results enhance our general understanding of sex determination strategies, which are already known to involve transcriptional regulation¹ and alternative RNA splicing^{2,3} in *Drosophila melanogaster*, DNA rearrangements in *Saccharomyces cerevisiae*⁴, and transcriptional regulation in mammals^{5,6}.

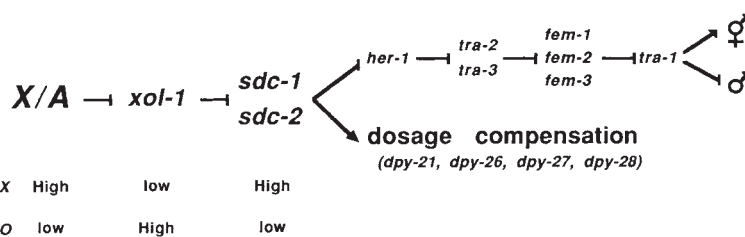
In *C. elegans*, the primary sex-determining signal, the X/A ratio (ratio of X chromosomes to autosomal sets), controls sex determination and dosage compensation through its effects on genes that coordinately control both processes^{7–12} (Fig. 1). Until now nothing was known regarding the molecular nature of the genes that respond most directly to the X/A ratio and act early in the regulatory hierarchy to control both sex determination and dosage compensation. The *sd-1* gene (sex determination and dosage compensation) is among the genes required in hermaphrodites (XX) to permit the correct choice of sexual fate and level of X chromosome expression. Mutations in *sd-1* have no effect in males (XO), but in XX animals result in masculinization (Tra phenotype), reflecting a defect in sex determination, and dumpy (Dpy), reflecting the elevated X-linked

Received 8 January; accepted 6 March 1991.

- Freissmuth, M., Casey, P. J. & Gilman, A. G. *FASEB J.* **3**, 2125–2131 (1989).
- Birnbaumer, L. A. *Rev. Pharmac. Tox.* **30**, 675–705 (1990).
- Bourne, H. R., Sanders, D. A. & McCormick, F. *Nature* **348**, 125–132 (1990).
- Masters, S. B. *et al.* *J. biol. Chem.* **264**, 15467–15474 (1989).

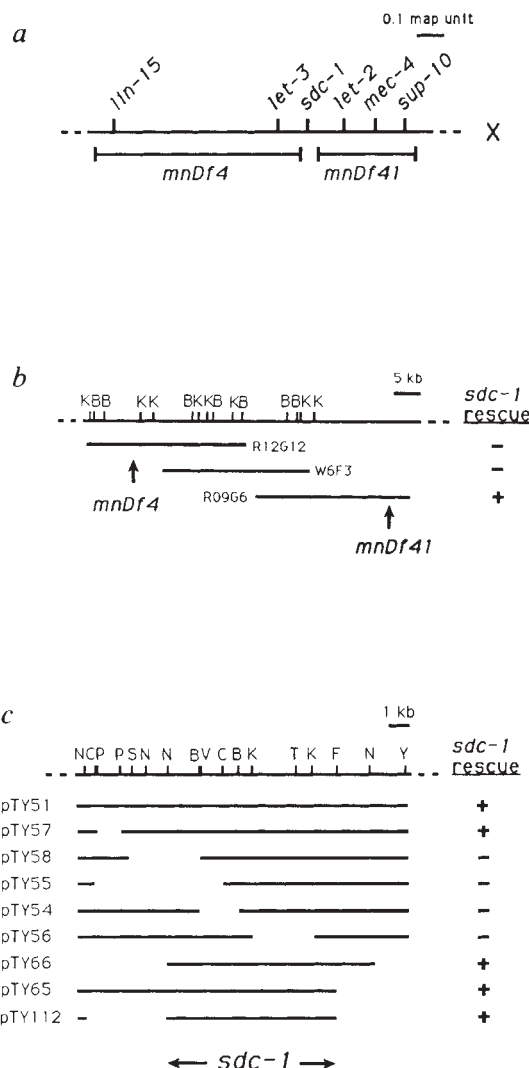
* To whom correspondence should be addressed.

FIG. 1 Regulatory network controlling *C. elegans* somatic sex determination and dosage compensation⁷. The genes *xol-1*, *sdca-1* and *sdca-2* are involved in the coordinate control of sex determination and dosage compensation. The branch of the pathway involved only in sex determination is composed of the *her-1* and *fem* genes, which are required for male sexual development⁸, and the *tra* genes, which are required for hermaphrodite development⁸. The genes that act downstream to implement dosage compensation include *dpy-21*, *dpy-26*, *dpy-27* and *dpy-28*, which are thought to reduce expression of each X chromosome in XX animals^{33,36}. Mutations in these genes result in overexpression of X-linked genes and a Dpy phenotype in XX animals; mutations in all but *dpy-21* result in a maternal-effect XX-specific lethality³³. The *sdca* genes control the hermaphrodite modes of sex determination and dosage compensation and act as negative regulators of the *her-1* sex determination gene and positive regulators of the XX-specific dosage compensation genes⁹⁻¹¹. Mutations in either *sdca-1* or *sdca-2* cause XX animals to adopt the male mode of both processes, resulting in masculinization and overexpression of X-linked genes. Unlike *sdca-2* mutations, *sdca-1* mutations exhibit strong maternal rescue but do not result in significant XX-specific lethality. The gene *xol-1* (for 'XO-lethal') is required in XO animals



to control the male modes of sex determination and dosage compensation. Mutations in *xol-1* cause XO animals to adopt the hermaphrodite modes of both processes, resulting in feminization, reduced X-linked transcript levels and XO-specific lethality. Mutations in *sdca-1* or *sdca-2* suppress the phenotypes caused by *xol-1* mutations, indicating that in XO animals the wild-type role of *xol-1* is to act as a negative regulator of the *sdca* genes¹². Arrows represent positive regulatory interactions; bars represent negative interactions. 'High' and 'low' represent the activity states of the various genes in hermaphrodites (XX) and males (XO).

FIG. 2 Genetic and physical maps of the *sdca-1* region on the X chromosome. a, A genetic map of the right arm of the *C. elegans* X chromosome¹⁰. The genetically determined extent of chromosomal deficiencies *mnDf4* and *mnDf41*, which flank *sdca-1*, are drawn below the map. The right breakpoint of *mnDf4* was previously positioned left of *let-3* (ref. 34), but an X chromosome carrying this deficiency does not complement *let-3* in our experiments. b, Localization of the *sdca-1* rescuing activity within an interval defined by deficiency breakpoints. A restriction map of a 65-kb genomic interval derived from restriction analysis of three overlapping cosmid clones is shown. The approximate positions of the right breakpoint of *mnDf4* and left breakpoint of *mnDf41* are indicated by arrows. Each cosmid in the interval was assayed by germ-line transformation for its ability to rescue the dumpy (Dpy), egg-laying (Egl) and sexually transformed (Tra) phenotypes of *sdca-1* mutants when incorporated into a stably transmitted extrachromosomal array. c, An expanded restriction map of the left half of the rescuing cosmid R09G6 and the structure of subclones of this region. Each subclone was tested for its ability to rescue an *sdca-1* mutant. The *sdca-1* rescuing activity resides in a 9-kb *NheI*-*SfiI* interval. Plasmid pTY51 is a R09G6 derivative that carries a 14-kb *BstEII* fragment deletion. The remainder of the subclones are pTY51 derivatives that contain deletions of the restriction fragments indicated by the open space. Plasmid pTY112 removes 4.1 kb of *NheI* genomic fragments from pTY65. Restriction site abbreviations are *Bam*HI (B), *Clal* (C), *Kpn*I (K), *Nhe*I (N), *Not*I (T), *Pst*I (P), *Pvu*II (V), *Sfi*I (F), *Stu*I (S), *BstEII* (Y). Southern hybridization experiments using probes that span the entire gene indicated that the 14 *sdca-1* alleles cause no obvious allele-specific polymorphisms. METHODS. To localize the deficiency breakpoints, cosmids at various distances from *sup-10* (positioned on the map by C. Cummins, J. Levin, P. Anderson & R. Horvitz, personal communication) were used to probe Southern blots of DNA derived from a wild-type N2 strain and strains homozygous for the *mnDf4* or *mnDf41* deficiency and heterozygous for duplication (*mnDp1/+*; *mnDf4* and *mnDp1/+*; *mnDf41*). Genomic fragments located outside the deficiency should be present in three copies, and those under the deficiency should be present in one copy. We localized the *mnDf4* right breakpoint to a 4-kb *Pst*I-*Stu*I fragment within cosmid R12G12 and the left breakpoint of *mnDf41* to a 3.5-kb *Hind*III fragment in R09G6. The position of the *sdca-1* gene was further localized by germ-line transformation rescue assays³⁵. Test DNA (20 µg ml⁻¹) and the dominant *rol-6(su1006)* roller marker (pRF4 plasmid at 100 µg ml⁻¹) were co-injected into the syncytial gonad of young adult *sdca-1* (*y67ts*) (ref. 10) animals grown at 15 °C. Injected animals were self-fertilized at 25 °C. F₁-transformed roller progeny carrying extrachromosomal arrays were scored for dumpy (Dpy), egg-laying defective (Egl) and sexually transformed (Tra) phenotypes, picked and self-fertilized. The F₂ broods were screened for the presence of roller animals. Most F₁ transformants are unstable and do not transmit the extrachromosomal arrays to their progeny. 'Stably' transformed lines were established from animals that transmitted the roller marked array to their progeny (5-80% transmission). A line was considered (+) in the rescue assay if both transient F₁ roller animals and stably transformed roller animals were rescued for Dpy, Egl and Tra phenotypes. A few clones (pTY54, pTY55, and pTY58) were capable of partially rescuing the *sdca-1* phenotype in the F₁ generation, but stable transformants carrying these plasmids were never rescued. Arrays of both pTY66 and pTY65 were shown to rescue *sdca-1*(*y36am*) (ref. 10). Plasmid pJS328 (gift of J. Shaw) contains a 2.2-kb *Bgl*II-*Hind*III fragment of



the *act-4* promoter and first intron inserted in pBluescript KS(+) and pTY59 contains a 330-bp *Bgl*II-*Eco*RI fragment of the *act-4* first intron inserted into pBluescript KS(+). The *act-4* first intron is the location of one of the X chromosome elements that feminizes animals with intermediate X/A ratios¹⁵. None of 50 *sdca-1* mutant animals transformed with pJS328 were suppressed; none of 240 *sdca-1* mutant animals transformed with pTY59 were suppressed.

transcript levels caused by a disruption of dosage compensation. Genetic experiments indicate that *sdsc-1* functions directly or indirectly to negatively regulate the *her-1* sex determination gene, which is required for male development, and to activate the downstream XX-specific genes required for proper dosage compensation^{9,10} (Fig. 1).

We have cloned *sdsc-1* using a physical map of the *C. elegans* genome consisting of groups of overlapping cosmid and yeast artificial chromosome clones (refs 13 and 14; A. Coulson and J. Sulston, personal communication). The *sdsc-1* gene is located on the right side of the X chromosome (Fig. 2a) in a region represented on the physical map by a series of overlapping cosmids spanning about 3 megabases, from the *sup-10* gene towards *sdsc-1*. Cosmid probes positioned on this physical map were used to search for the breakpoints of two deficiencies that flank *sdsc-1* (Fig. 2a and b). The deficiency breakpoints localized the *sdsc-1* gene to a 65-kilobase (kb) interval (Fig. 2b).

To confirm the position of the *sdsc-1* gene in this region, we used germ-line transformation to rescue an *sdsc-1* temperature-sensitive mutant. Of the three cosmids that span the *sdsc-1* region, only a single cosmid, R09G6, was capable of rescuing the *sdsc-1* mutant phenotypes (Fig. 2b). A series of R09G6 subclones was constructed and tested for the ability to rescue the *sdsc-1* mutant phenotypes (Fig. 2c). These experiments localized the *sdsc-1* rescuing activity to a 9-kb *NheI-SfiI* genomic DNA fragment. This fragment also rescues the mutant phenotype of animals carrying a putative null amber allele of *sdsc-1*.

Genetic experiments indicate that large X chromosome duplications that do not cover *sdsc-1* can feminize *sdsc-1* mutant animals and thus suppress the *sdsc-1* phenotypes¹⁰. We thus had to exclude the possibility that the extrachromosomal arrays generated in our transformation experiments might mimic this effect rather than reflect the isolation of the *sdsc-1* gene. Neither adjacent cosmids nor two clones, containing X chromosome sequences, previously shown to act as feminizing elements in transformation experiments¹⁵, can rescue the *sdsc-1* mutant phenotypes (Fig. 2).

To characterize transcripts derived from the *sdsc-1* region, a 17-kb genomic clone containing *sdsc-1* rescuing activity (pTY51) was subdivided into six fragments that were used individually to probe poly(A)⁺ RNA isolated from a wild-type hermaphrodite strain and from a male-producing *him-8* strain¹⁶ (not shown). We identified a single 3.8-kb transcript that spans most of the 9-kb *sdsc-1* interval. Analysis of the developmental time course of the *sdsc-1* transcript showed that the transcript is preferentially expressed during the embryonic period (Fig. 3). The *sdsc-1* transcript levels decrease as development proceeds through the larval stages and then increase in adults (Fig. 3). This developmental expression pattern is consistent with temperature-shift experiments that demonstrate a requirement for *sdsc-1* during the first half of embryogenesis¹⁰. It is also consistent with the *sdsc-1* phenotype being maternally rescuable, indicating that *sdsc-1* should be expressed in the adult hermaphrodite germ line⁹.

DNA sequence analysis revealed that *sdsc-1* encodes a 1,203-amino-acid protein containing seven zinc-finger motifs (Fig. 4a). Although five of the motifs in *sdsc-1* resemble the consensus C₂H₂ TFIIIA motif^{17,18} (Fig. 4b), two show more variability than those in most other proteins containing many zinc-fingers^{17,19-21}. The TFIIIA zinc-finger DNA-binding motifs consist of conservatively positioned pairs of cysteine and histidine residues that coordinate a metal ion (usually zinc)^{17,18}. The *sdsc-1* zinc finger motifs that resemble the consensus TFIIIA motif are interspersed with those that have the histidine pair replaced with a cysteine-histidine pair. This type of variation is observed in some zinc-finger proteins known to bind DNA, including the major histocompatibility complex enhancer binding protein, MBP-1 (ref. 22), the SW15 protein involved in yeast mating-type switching²³ and the PRDII-BF1 protein that positively regulates the human interferon β gene²⁴. In contrast to most multiple

zinc-finger proteins^{17,19-21}, the *sdsc-1* fingers are not arranged in a regularly spaced tandem array, but rather in four widely spaced clusters of single and paired motifs (Fig. 4c). In this regard *sdsc-1* is reminiscent of MBP-1 and PRDII-BF1, which have structurally distinct zinc-fingers positioned in the amino- and carboxy-terminal regions of their protein products. In light of the similarity between *sdsc-1* and known DNA binding proteins, it seems likely that *sdsc-1* functions as a maternally supplied, embryonic transcription factor. However, we cannot exclude the interesting possibility of *sdsc-1* acting primarily as an RNA-binding protein such as the p43 zinc-finger protein that preferentially binds 5S RNA in *Xenopus*²⁵.

Genetic evidence has already indicated that *sdsc-1*, together with *sdsc-2* and the newly identified *sdsc-3* gene^{26,27} (not shown in Fig. 1), promotes hermaphrodite sexual development by negatively regulating *her-1*, a sex determination gene required for male development^{10,11}. Molecular confirmation has come from the observation that *her-1* transcript levels, normally very low in hermaphrodites but high in males^{28,29}, are greatly elevated in XX *sdsc-1*, *sdsc-2* (ref. 28) and *sdsc-3* (L. DeLong, J. Plenefisch, R. Klein and B.J.M., manuscript in preparation) mutant animals. These data, together with the identification of an *sdsc-1* zinc-

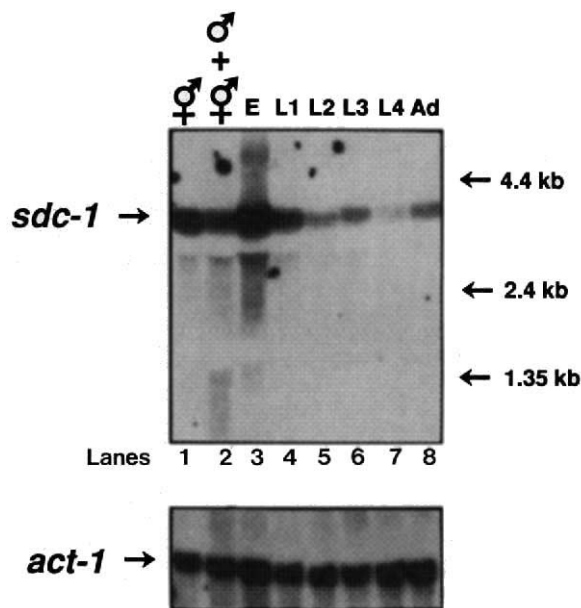


FIG. 3 Developmental regulation of *sdsc-1* transcript levels. A northern blot of poly(A)⁺ RNAs (10 µg per lane) from different developmental stages of the *C. elegans* life cycle. Lanes 1 and 2, RNA isolated from mixed-staged cultures of wild-type N2 strain (♂) or a male-producing *him-8* strain (♂ and ♀, approximately 30% males). Lanes 3–8, RNA isolated from synchronized hermaphrodite cultures grown to different developmental stages—eggs (E), the four larval stages (L1–L4), and adults before the onset of egg production (Ad). The northern blot was probed individually with a 3.0-kb *EcoRI* fragment from complementary DNA clone λTY1, and with an *act-1* specific actin clone (pW-16-210; gift of M. Krause) to control for the amount of mRNA in each lane. RNA size markers appear on the right of the autoradiograph. Mixed-stage RNA from the *him-8* strain does not reveal any alternative, male-specific transcript from the *sdsc-1* region.

METHODS. Nematodes were grown in liquid culture and synchronized as described in ref. 36. Synchronized L1 larval populations were fed and L1 larvae collected after 2 h at 20 °C. L2 (26 h), L3 (34 h), L4 (44 h) and young adults (50 h) were collected at later times after confirming the staging by microscopic examination of animals using Nomarski optics. To isolate RNA, frozen packed nematodes (5 ml) were crushed under liquid nitrogen with a mortar and pestle, transferred to a tube containing 5 ml phenol, 5 ml CHCl₃ and 15 ml 1% SDS 1% β -mercaptoethanol, 10 mM EDTA, 0.2 M Tris-Cl, pH 7.5, and 0.5 M NaCl. The mixture was sonicated for 1 min in six bursts using a microtip at maximum power, and the phases separated by centrifugation. After four subsequent phenol extractions, the aqueous layer was loaded onto an oligo d(T) column, and poly(A)⁺ RNA was isolated as described in ref. 37. Electrophoresis, blotting and probing of northern blots were as described in ref. 37.

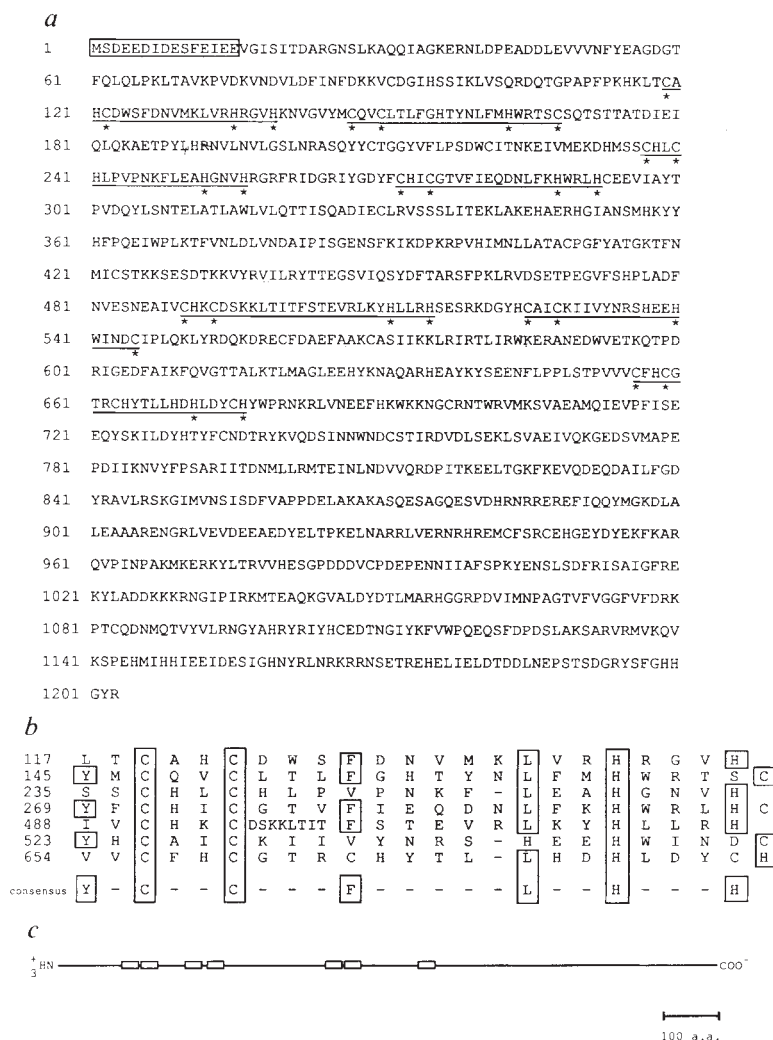


FIG. 4 **a**, The deduced amino-acid sequence of *sdc-1* (single-letter code). A highly acidic region of -9 charge is boxed. Each of the zinc fingers is underlined, and the relevant cysteines and histidines are indicated by stars. The GenBank-EMBL accession number for the *sdc-1* DNA sequence is X58520. **b**, The sequence of the seven zinc-finger motifs found in *sdc-1*. The numerical position of the initial amino acid of each zinc-finger is located to the left of the amino-acid sequences. The TFIIIA-like zinc-finger motif consensus sequence¹⁷ is listed below; *sdc-1* amino acids matching the consensus are boxed. **c**, A schematic of the *sdc-1* protein highlighting (boxes) the position of the seven zinc-finger motifs within the protein.

METHODS. A mixture of two probes from the left and right side of pTY66 was used to isolate three independent *sdc-1* cDNA clones from a mixed stage hermaphrodite library (gift of S. Kim). The cDNAs (λ TY1, 3.6 kb; λ TY7, 2.8 kb; and λ TY3, 2.2 kb) were shown by restriction analysis and partial sequence analysis to contain different sized cDNA clones of the same mature message. Further, we demonstrated that these three cDNAs hybridize only to genomic sequences corresponding to the *sdc-1* gene as defined by transgenic rescue. Because the largest cDNA was incomplete, the clone of the 5' end of *sdc-1* message was isolated by polymerase chain reaction (PCR) techniques. The sequence of *sdc-1* message was compiled from sequencing of the λ TY1 cDNA insert, genomic sequences 5' of this cDNA, and 5' end PCR products. In addition, genomic sequences corresponding to at least 90% of the cDNA sequences were determined and found to be identical. Many transcripts in *C. elegans* are *trans*-spliced to one of two leader RNAs³⁸. We amplified by PCR and sequenced the 5' end of the *sdc-1* message using an oligonucleotide of SL1 *trans*-splice leader (5'-TCTAGAATTCGCGGTTTAATTACCCAAG-TTG-3') and a complementary oligonucleotide (5'-AAATCCACTG-GTTTCACAGCTG-3') from the 5' end of our largest cDNA clone. Mature *trans*-spliced message is 3.8 kb without the poly(A) tail. But primer extension of poly(A)⁺ egg RNA using an oligonucleotide 54 bp 3' of the *trans*-splicing site (5'-CCTCTTCGATTTCAAGCTT-TCATC-3') suggests that most (>90%) of egg *sdc-1* RNA is not *trans*-spliced and probably initiates in a cluster 70–80 bp upstream of the *trans*-splicing site.

finger motif, suggest that a component of the early response to the X/A ratio involves transcriptional regulation. But the sexual transformation of XX animals towards the male fate is not as complete in *sdc-1* mutants as it is in *sdc-2* or *sdc-3* mutants, suggesting that *sdc-1* does not act alone in regulating the downstream sex determination genes⁷. A similar situation arises in the control of dosage compensation. Ultimately, *sdc-1* in combination with other *sdc* genes must activate the downstream hermaphrodite-specific genes that equalize X-chromosome expression between the sexes.

If *sdc-1* is a transcriptional regulator, then its target(s) might either be other *sdc* genes or the downstream genes that specifically control sex determination or dosage compensation. In the latter case, *sdc-1* would function simultaneously as a negative regulator of sex determination genes and a positive activator of dosage compensation genes. In this regard, the *sdc-1* protein might be similar to the glucocorticoid receptor³⁰ or the yeast MCM1 protein. The MCM1 protein interacts with the yeast mating-type protein MAT α 1 to activate transcription of α -specific genes in α cells³¹, but interacts with the yeast mating-type protein MAT α 2 to repress transcription of α -specific genes in α cells³². Delineating the *sdc-1* targets and determining the functional roles of the zinc fingers will greatly improve our understanding of the mechanisms underlying these complex developmental decisions. □

Received 11 January; accepted 1 March 1991.

1. Erickson J. W. & Cline, T. W. *Science* **251**, 1071–1074 (1991).
2. Bell, L. R., Maine, E. M., Schedl, P. & Cline, T. W. *Cell* **55**, 1037–1046 (1988).
3. Baker, B. S. *Nature* **340**, 521–524 (1989).

4. Herskowitz, I. *Microbiol. Rev.* **52**, 536–553 (1988).
5. Sinclair, A. H. et al. *Nature* **346**, 240–244 (1990).
6. Gubbay, J. et al. *Nature* **346**, 245–250 (1990).
7. Villeneuve, A. M. & Meyer, B. J. *Adv. Genet.* **27**, 117–188 (Academic, San Diego, 1990).
8. Hodgkin, J. *Nature* **344**, 721–728 (1990).
9. Villeneuve, A. M. & Meyer, B. J. *Cell* **48**, 25–37 (1987).
10. Villeneuve, A. M. & Meyer, B. J. *Genetics* **130**, 91–144 (1990).
11. Nusbaum, C. & Meyer, B. J. *Genetics* **122**, 579–593 (1989).
12. Miller, L. M., Plenefisch, J. D., Casson, L. P. & Meyer, B. J. *Cell* **55**, 167–183 (1988).
13. Coulson, A., Sulston, J., Brenner, S. & Karn, J. *Proc. natn. Acad. Sci. U.S.A.* **83**, 7821–7825 (1986).
14. Coulson, A. et al. *Nature* **335**, 184–186 (1990).
15. McCoubrey, W. K., Nordstrom, K. D. & Meneely, P. M. *Science* **242**, 1146–1151 (1988).
16. Hodgkin, J., Horvitz, H. R. & Brenner, S. *Genetics* **91**, 67–94 (1979).
17. Klug, A. & Rhodes, D. *Trends Biochem. Sci.* **12**, 464–469 (1987).
18. Evans, R. M. & Hollenberg, S. M. *Cell* **52**, 1–3 (1988).
19. Miller, J., McLachlan, A. D. & Klug, A. *EMBO J.* **41**, 1609–1614 (1985).
20. Chowdhury, K., Deutsch, U. & Gruss, P. *Cell* **48**, 771–778 (1987).
21. Page, D. et al. *Cell* **51**, 1091–1104 (1987).
22. Baldwin, A. S. Jr, LeClair, K. P., Singh, H. & Sharp, P. A. *Molec. cell. Biol.* **10**, 1406–1414 (1990).
23. Nagai, K., Nakaseko, Y., Nasmyth, K. & Rhodes, D. *Nature* **332**, 284–286 (1988).
24. Fan, C.-M. & Maniatis, T. *Genes Dev.* **4**, 29–42 (1990).
25. Joho, K. E., Darby, M. K., Crawford, E. T. & Brown, D. D. *Cell* **61**, 293–300 (1990).
26. DeLong, L. thesis, Massachusetts Inst. Technol. (1990).
27. Plenefisch, J. D. thesis, Massachusetts Inst. Technol. (1990).
28. Trent, C. et al. *Mechanisms Dev.* (in the press).
29. Schauer, I. E. & Wood, W. B. *Development* **110**, 1303–1317 (1990).
30. Diamond, M. I., Miner, J. M., Yoshinaga, S. K. & Yamamoto, K. R. *Science* **249**, 1266–1272 (1990).
31. Bender, A. & Sprague, G. F. *Cell* **50**, 681–691 (1987).
32. Keleher, C. A., Goutte, C. & Johnson, A. D. *Cell* **53**, 927–936 (1988).
33. Plenefisch, J. D., DeLong, L. & Meyer, B. J. *Genetics* **121**, 57–76 (1989).
34. Meneely, P. M. & Herman, R. K. *Genetics* **92**, 99–115 (1979).
35. Mello, C. thesis, Harvard Univ. (1990).
36. Meyer, B. J. & Casson, L. P. *Cell* **47**, 871–881 (1986).
37. Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).
38. Huang, X.-Y. & Hirsh, D. *Proc. natn. Acad. Sci. U.S.A.* **86**, 8640–8644 (1989).

ACKNOWLEDGEMENTS. We thank T. Cline, J. Erickson, R. Tjian and all the members of the Meyer laboratory for their comments. A. Coulson, J. Sulston for cosmid clones, and C. Mello and J. Kramer for the pRF4 plasmid. Some of the nematode strains were provided by the *Caenorhabditis* Genetics Center, which is funded by the National Center for Research Resources. This work was supported by the NSF and the US Public Health Service. M.L.N. is a fellow of the Helen Hay Whitney Foundation.

Cloning of a complementary DNA for a protein-tyrosine kinase that specifically phosphorylates a negative regulatory site of p60^{c-src}

Shigeyuki Nada, Masato Okada*, Alasdair MacAuley†, Jonathan A. Cooper† & Hachiro Nakagawa

Division of Protein Metabolism, Institute for Protein Research, Osaka University, 3-2 Yamadaoka, Suita, Osaka 565, Japan

† Fred Hutchinson Cancer Research Center, Seattle, Washington 98104, USA

* To whom correspondence should be addressed

THE protein-tyrosine kinase activity of the proto-oncogene product p60^{c-src} is negatively regulated by the phosphorylation of a tyrosine residue close to the C terminus, tyrosine 527 (refs 1–11). The phosphorylation might be catalysed by a so-far-unidentified tyrosine kinase, distinct from p60^{c-src} (ref. 7). Recently we purified a protein-tyrosine kinase that specifically phosphorylates tyrosine 527 of p60^{c-src} from neonatal rat brain^{8,12,13}. We have now confirmed the specificity of this enzyme by using a mutant p60^{c-src} that has a phenylalanine instead of tyrosine 527, and cloned a complementary DNA that encodes the enzyme. The enzyme is similar to kinases of the *src* family in that it has two conserved regions, *Src*-homology regions 2 and 3, upstream of a tyrosine kinase domain. The amino-acid identity of each region is no more than 47%, however, and the enzyme lacks phosphorylation sites corresponding to tyrosines 416 and 527 of p60^{c-src} and has no myristylation signal. These results suggest that this protein-tyrosine kinase, which might negatively regulate p60^{c-src}, represents a new type of tyrosine kinase.

The specificity of our novel tyrosine kinase (termed 'c-Src

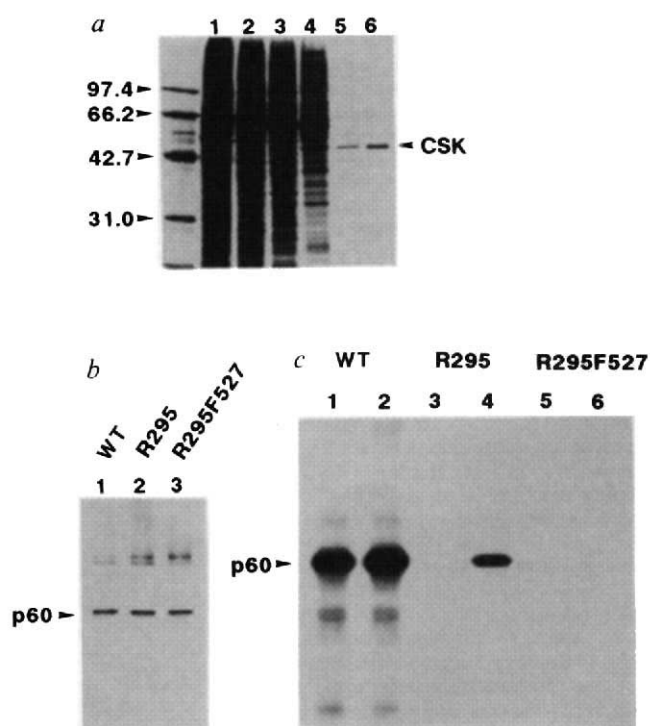
kinase' or 'CSK') was further confirmed by an experiment using kinase-inactive mutants of p60^{c-src}. Wild-type p60^{c-src} (p60^{WT}) and two kinds of mutants (p60^{R295} and p60^{R295F527}) were expressed in yeast cells^{11,14}. Mutant p60^{R295} was generated by introducing a mutation into the codon for lysine 295 so that it specified arginine; p60^{R295F527} was generated from p60^{R295} by introducing a further mutation in the codon for Tyr 527 so that it specified phenylalanine. When synthesized in yeast, p60^{WT} is phosphorylated to low stoichiometry at Tyr 527 only¹⁴, and p60^{R295} is not phosphorylated at all¹¹, so these molecules should be substrates for kinases that are specific for Tyr 527. These p60^{c-src} derivatives were purified from yeast cell lysates with a monoclonal antibody, which does not interfere with phosphorylation by CSK. Immunoprecipitates containing similar quantities of the three proteins (Fig. 1b) were incubated with highly purified CSK (Fig. 1a). The phosphorylation of p60^{WT} was greatly increased by the addition of CSK (Fig. 1c, lanes 1 and 2). Mutant p60^{R295} was also phosphorylated (lanes 3 and 4), whereas p60^{R295F527} was not phosphorylated at all (lanes 5 and 6). These results imply that CSK is highly specific for Tyr 527, and show that CSK does not redirect the autophosphorylation activity of p60^{c-src} to Tyr 527. When these p60^{c-src} mutants were incubated with purified insulin receptor, a tyrosine kinase with broad substrate specificity, neither p60^{R295} nor p60^{R295F527} was markedly phosphorylated (data not shown). As specific phosphorylation of Tyr 527 of p60^{c-src} by any other kinase has not been reported, CSK is a strong candidate for the specific kinase involved in the regulation of p60^{c-src}.

To characterize the enzyme further, we cloned the cDNA that encodes CSK and deduced its primary structure (Fig. 2). The complete nucleotide sequence of the cDNA is shown in Fig. 2. The cDNA insert has a single open reading frame which encodes a protein of 450 amino-acid residues with a relative molecular mass of 50,753.

To show that the cDNA clone encodes CSK, we transfected the cDNA into yeast cells carrying the mutant p60^{c-src} proteins (Fig. 3a), and examined whether the cDNA product could

FIG. 1 Phosphorylation of mutant p60^{c-src} by CSK. **a**, Sample of each purification step of CSK was analysed by SDS-PAGE and staining with silver. Lane 1, Nonidet P-40 extract; lane 2, DEAE-cellulose column chromatography; lane 3, poly(Glu, Tyr) Sepharose CL-4B; lane 4, Mono Q; lane 5, Sephacryl S200HR; lane 6, Mono S. Positions of relative molecular mass (*M_r*) markers (in *M_r* × 10⁻³) are indicated. **b**, Expression of p60^{WT} (lane 1), p60^{R295} (lane 2) and p60^{R295F527} (lane 3) in yeast cells was detected by immunoblotting of the immunoprecipitates. The bands above p60 are immunoglobulin. **c**, Phosphorylation of p60^{c-src} by CSK was determined by incubating immunoprecipitates of p60^{WT} (lanes 1 and 2), p60^{R295} (lanes 3 and 4) and p60^{R295F527} (lanes 5 and 6) with (lanes 2, 4 and 6) or without (lanes 1, 3 and 5) CSK. The reaction products were analysed by SDS-PAGE and autoradiography.

METHODS. **a**, CSK was purified from the membrane fraction of neonatal rat brain by sequential column chromatography. The method was essentially the same as that described previously⁸, but with changes to improve the yield and purity. In brief, after DEAE-cellulose column chromatography, the sample was applied to a poly(Glu, Tyr) Sepharose CL-4B column, and eluted with 0.1 M NaCl. The eluate was applied to a Mono Q column, and eluted with a linear gradient of 0–0.35 M NaCl. The materials eluted at ~0.25 M NaCl were separated on a Sephacryl S200HR column. Finally, the active fractions were applied to a Mono S column, and eluted with a linear gradient of 0–0.2 M NaCl. CSK was eluted at about 0.1 M NaCl. By this improved method, we obtained about 100 µg highly purified CSK from 300 g neonatal rat brains. **b**, Yeast cells expressing p60^{WT} (pMX), p60^{R295} (JC851) and p60^{R295F527} (JC861) were generated as described previously^{11,13}. Cell culture, lysis, immunoprecipitation with monoclonal antibody 327 and immunoblotting with the same antibody were carried out by the method of Cooper and Runge¹⁴, except that the antigens were detected with peroxidase-antiperoxidase complex. In this experiment, immunoprecipitates obtained from 1 × 10⁹ yeast cells were analysed. **c**, Phosphorylation by CSK was determined as described previously⁸. The reaction mixture (10 µl) contained 50 mM Tris-HCl, pH 7.4, 3 mM MnCl₂, 0.1 mM Na₂VO₄, 1 µM [γ-³²P]ATP (74 kBq), immunoprecipitate of p60^{c-src} obtained from 1 × 10⁹ yeast cells and 50 ng of CSK. After incubation for 10 min at 30°C, the reaction was

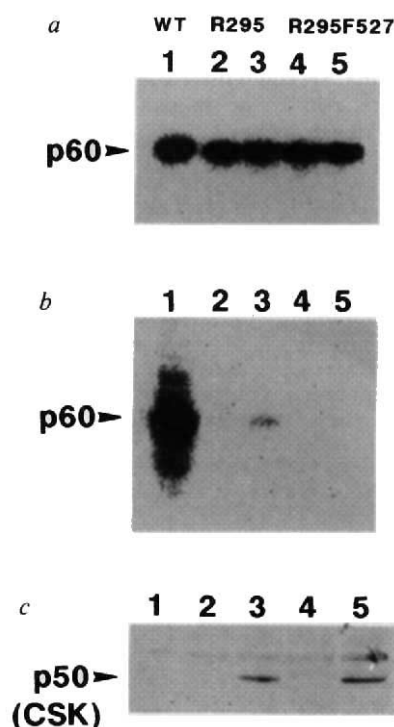


terminated by adding 10 µl of SDS sample buffer and boiling for 2 min. The sample was then subjected to SDS-PAGE, and the phosphoproteins were located by autoradiography.

NATURE • VOL 351 • 2 MAY 1991

FIG. 3 Phosphorylation of Tyr 527 by the cDNA product in yeast cells. The cDNA was transfected into yeast cells carrying the mutant $p60^{c-src}$ proteins, and their tyrosine phosphorylation was estimated by immunoblot analysis with anti-phosphotyrosine antibody. **a**, Expression of $p60^{c-src}$ in yeast cells carrying $p60^{WT}$ (lane 1), $p60^{R295}$ (lane 2), $p60^{R295F527}$ and pHMCSK (lane 3), $p60^{R295F527}$ (lane 4), and $p60^{R295F527}$ and pHMCSK (lane 5) was confirmed by immunoblotting with monoclonal antibody 327. **b**, Phosphorylation of $p60^{c-src}$ in the same cultures was detected by immunoblotting with anti-phosphotyrosine antibody. **c**, Expression of the CSK cDNA product (lanes 3 and 5, p50) was detected by immunoblotting with an antibody raised against bacterially expressed protein.

METHODS. A fragment containing the open reading frame (nucleotides 71–1,452, Fig. 2) was amplified by PCR using two primers attached to *Sall* linkers, then the *BglII*–*XhoI* fragment (71–1,405) was exchanged for the same fragment of original clone to exclude the possibility of mutations occurring during PCR. The residual fragment (1,406–1,452) was confirmed not to contain any mutations by sequencing the fragment. The *BglII*–*Sall* fragment obtained was ligated to *BamHI*–*Sall* fragment of pHM209, a derivative of pHM153³² containing the *TRP1* gene. The expression plasmid (pHMCSK) was transfected into the yeast strain expressing $p60^{R295}$ (JC851) and $p60^{R295F527}$ (JC861). *Trp*⁺ and *Ura*⁺ transformants were selected, and grown as described previously^{11,13}. Yeast cells expressing $p60^{WT}$ (lane 1), $p60^{R295}$ (lane 2), $p60^{R295}$ and pHMCSK (lane 3), $p60^{R295F527}$ (lane 4), and $p60^{R295F527}$ and pHMCSK (lane 5) were collected, and immediately lysed in SDS-PAGE sample buffer by boiling. Each sample (equivalent to 2×10^8 cells) was then immunoblotted. Detection of $p60^{c-src}$ was with monoclonal antibody 327 (**a**), and phosphotyrosine was detected with monoclonal antibody PY20 (ICN) (**b**). In this experiment, these monoclonal antibodies were visualized with ¹²⁵I-labelled anti-mouse IgG (sheep). The radioactivity bound was measured using Biomage-Analyzer BAS2000 (Fuji Film). Expression of the CSK cDNA product was detected with antibody raised against bacterially expressed protein (**c**). BluescriptII SK+ containing the *BglII*–*Sall* fragment was digested with *XbaI*, filled with Klenow fragment, and ligated to allow in-frame translation. The expression plasmid was transfected into *Escherichia coli* (TG1). The lacZ fusion protein was induced by IPTG, and



purified from inclusion bodies by preparative SDS-PAGE. Antiserum was obtained from a rabbit immunized with the purified protein. The rabbit polyclonal antibody was detected with peroxidase–antiperoxidase complex.

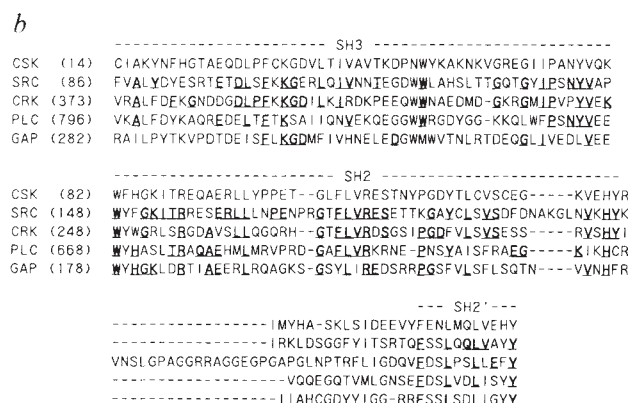
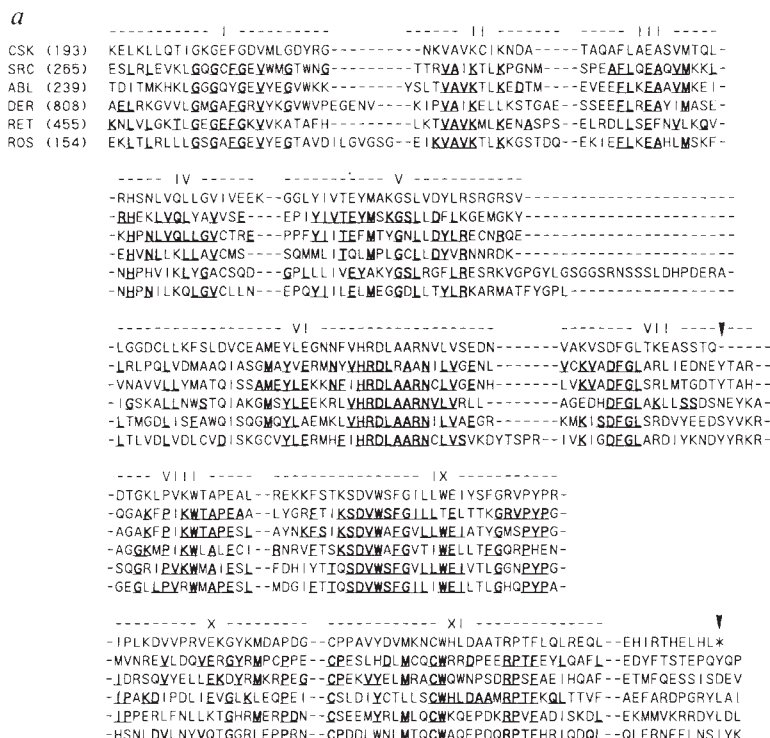


FIG. 4 **a**, Amino-acid alignment of the conserved kinase domain in CSK; *c-src* gene product (SRC; a representative of the *src* subfamily)¹⁶; *c-abl* gene product (ABL; *abl* subfamily)¹⁷; *Drosophila* gene product related to the EGF receptor (DER, EGF receptor subfamily)¹⁸; cellular oncogene product activated by recombination (RET, PDGF receptor subfamily)¹⁹ and *c-ros* gene product (ROS; insulin receptor subfamily)²⁰. Underlined bold type indicates residues identical to the kinase domain sequence in CSK. Roman numbers designate subdomains conserved in protein kinases¹⁵. The number in parentheses is the first amino acid of sequence shown. Sequence identities are 46% (SRC), 46% (ABL), 41% (DER), 39% (RET) and 41% (ROS). Sequence identities to

other *src*-family kinases are 45% (*fgr*), 44% (*lyn*), 43% (*yes*), 43% (*hck*), 42% (*fyn*) and 42% (*lck*)¹⁵. The percentage identities were determined over the length of the kinase domain of CSK; residues in insertions were ignored for these calculations. Arrows denote the residues corresponding to Tyr 416 and Tyr 527 of $p60^{c-src}$. **b**, Amino-acid alignment of SH2, SH2' and SH3 regions in CSK, *c-src* gene product (SRC)¹⁶, *gag-crk* gene product (CRK)²⁴, phospholipase C- γ (PLC)²⁵ and GTPase activator protein (GAP)²⁶. Total sequence identities in these regions are 47% (SRC), 40% (CRK), 30% (PLC) and 27% (GAP). The percentage identities were determined as described above.

important in the regulation of tyrosine kinase and in the interaction with cellular proteins that serve as substrates or regulators. These regions in CSK could therefore be involved in recognition of its substrate, p60^{src}, or interaction with a regulator that modulates its kinase activity.

CSK lacks a myristylation signal at the N terminus and has no potential membrane-spanning sequences, suggesting that it is a cytosolic protein. Although CSK was originally purified from membrane fractions, immunoblotting with anti-CSK antibody indicates that CSK is also present in the soluble fraction (data not shown). This is consistent with a soluble mutant of p60^{src} being also phosphorylated effectively at Tyr 527²⁹. CSK recovered from membrane fractions might have been bound to p60^{src}.

Northern blot analysis using a probe for CSK mRNA identified a 2.4-kilobase (kb) transcript which was present in all adult rat tissues examined, but markedly concentrated in neonatal brain (data not shown). Because p60^{src} is concentrated in developing rat brain³⁰, CSK could have an important role in the regulation of p60^{src} during neuronal differentiation. In thymus and spleen, cross-hybridizing transcripts of 2.2 kb were detected at a similar level to that of the neonatal transcript. This raises the possibility that a subtype of CSK exists in these tissues, which indeed contain different types of *src*-family kinases, such as the gene products of *lck*, *fyn* and *lyn*. These kinases might be involved in the regulation of lymphoid-cell differentiation³¹. Therefore, distinct subtypes of CSK may regulate *src*-family kinases in some non-neuronal cell types. □

Received 19 October 1990; accepted 5 March 1991.

- Cooper, J. A., Gould, K. L., Cartwright, C. A. & Hunter, T. *Science* **231**, 1431–1434 (1986).
- Cooper, J. A. & King, C. S. *Molec. cell. Biol.* **6**, 4467–4477 (1986).
- Courtneidge, S. A. *EMBO J.* **4**, 1471–1477 (1985).
- Kimieck, T. E. & Shalloway, D. *Cell* **49**, 65–73 (1987).
- Piwnicka-Worms, H., Saunders, K. B., Roberts, T. M., Smith, A. E. & Cheng, S. H. *Cell* **49**, 75–82 (1987).
- Cartwright, C. A., Eckhart, W., Simon, S. & Kaplan, P. L. *Cell* **49**, 83–91 (1987).
- Jove, R., Kornbluth, S. & Hanafusa, H. *Cell* **50**, 937–943 (1987).
- Okada, M. & Nakagawa, H. *J. Biol. Chem.* **264**, 20886–20893 (1989).
- Hunter, T. & Sefton, B. M. *Proc. natn. Acad. Sci. U.S.A.* **77**, 1311–1315 (1980).
- Kornbluth, S., Jove, R. & Hanafusa, H. *Proc. natn. Acad. Sci. U.S.A.* **84**, 4455–4459 (1987).
- Cooper, J. A. & MacAuley, A. *Proc. natn. Acad. Sci. U.S.A.* **85**, 4232–4236 (1988).
- Okada, M. & Nakagawa, H. *Biochem. biophys. Res. Commun.* **154**, 796–802 (1988).
- Okada, M. & Nakagawa, H. *J. Biochem.* **104**, 297–305 (1988).
- Cooper, J. A. & Runge, K. *Oncogene Res.* **1**, 297–310 (1987).
- Hanks, S. K., Quinn, A. M. & Hunter, T. *Science* **241**, 42–52 (1988).
- Takeya, T. & Hanafusa, H. *Cell* **32**, 881–890 (1983).
- Shitelman, E., Lifshitz, B., Gale, R. P., Roe, B. A. & Canaan, E. *Cell* **47**, 277–284 (1986).
- Livneh, E., Glazer, L., Segal, D., Schlessinger, J. & Shilo, B.-Z. *Cell* **40**, 599–607 (1985).
- Takahashi, M. & Cooper, G. M. *Molec. cell. Biol.* **7**, 1378–1385 (1987).

- Matsushima, H., Wang, L.-H. & Shibuya, M. *Molec. cell. Biol.* **6**, 3000–3004 (1986).
- Hunter, T. *Cell* **49**, 1–4 (1987).
- Sadowski, I., Stone, J. C. & Pawson, T. *Molec. cell. Biol.* **6**, 4396–4408 (1986).
- Hirai, H. & Varmus, H. E. *Molec. cell. Biol.* **10**, 1307–1318 (1990).
- Mayer, B. J., Hamaguchi, M. & Hanafusa, H. *Nature* **332**, 272–275 (1988).
- Stahl, M. L., Ferenz, C. R., Kelleher, K. L., Kriz, R. W. & Knopf, J. L. *Nature* **332**, 269–272 (1988).
- Vogel, U. S. *et al. Nature* **335**, 90–93 (1988).
- Meisenhelder, J., Suh, P.-G., Rhee, S. G. & Hunter, T. *Cell* **57**, 1109–1122 (1989).
- Molloy, C. J. *et al. Nature* **342**, 711–714 (1989).
- Schuh, S. M. & Brugge, J. S. *Molec. cell. Biol.* **8**, 2465–2471 (1988).
- Cartwright, C. A., Simantov, R., Cowan, W. M., Hunter, T. & Eckhart, W. *Proc. natn. Acad. Sci. U.S.A.* **85**, 3348–3352 (1988).
- Veillette, A., Bookman, M. A., Horak, E. M., Samelson, L. E. & Bolen, J. B. *Nature* **338**, 257–259 (1988).
- Matsuzaki, H., Nakajima, R., Nishiyama, J., Araki, H. & Oshima, Y. *J. Bact.* **172**, 610–618 (1990).

ACKNOWLEDGEMENTS. We thank S. Tsunasawa (Institute for Protein Research, Osaka University) for peptide sequence analysis. This work was supported by the Ministry of Education, Science and Culture of Japan, the US Public Health Service and the Yamanouchi Foundation for Research on Metabolic Disorders. Sequence data submitted to EMBL Data Library, accession number X58631.

Specific binding of antigenic peptides to cell-associated MHC class I molecules

Immanuel F. Luescher, Pedro Romero, Jean-Charles Cerottini & Janet L. Maryanski

Ludwig Institute for Cancer Research, Lausanne Branch, 1066 Epalinges, Switzerland

T LYMPHOCYTES recognize antigen in the form of peptides that associate with specific alleles of class I or class II major histocompatibility (MHC) molecules^{1,2}. By contrast with the clear MHC allele-specific binding of peptides to purified class II molecules^{3–6} purified solubilized class I molecules either bind relatively poorly⁷ or show degenerate specificity^{8–11}. Using photoaffinity labelling, we demonstrate here the specific interaction of peptides with cell-associated MHC class I molecules and show that this involves metabolically active processes.

The relative efficiency of the interaction of different peptides with the K^d molecule was assessed in a functional competition assay^{12,13} using a K^d-restricted cytotoxic T lymphocyte (CTL) clone (Fig. 1a). Of the peptides recognized by K^d-restricted CTL, one of the most potent competitors was an octapeptide from the mouse malaria *P. berghei* circumsporozoite protein, CS P.b. 253–260 (ref. 14, and P.R. *et al.*, manuscript submitted) (Fig. 1a). The K^d-restricted peptide P198^{14–24} (ref. 15) and a polyproline-containing peptide analogue (AYP₅TLA) (ref. 16) also competed, but the L^d-restricted peptides P91A^{12–24} (ref. 17) and LCMV NP 118–132 (ref. 18), and the D^b-restricted peptide Ad5 E1a 234–243 (ref. 19) failed to compete efficiently. The CS P.b. peptide was still an efficient competitor after N-terminal conjugation with the photoreactive iodo-4-azidosalicyl-

oyl (IASA) group (Fig. 1b); the IASA group itself was not responsible for the competitive activity of the conjugates because the derivative having a deletion at Tyr 253 (IASA conjugated with the sequence IPSAEKI (one-letter amino-acid code) was less active by at least 100-fold compared with IASA-YIPSAEKI (Fig. 1b). Moreover, the IASA-YIPSAEKI derivative was recognized by some CS-specific CTL clones (not shown). Efficient competitor activity, but not recognition, was retained after biotinylation of the Lys 259 residue (Fig. 1b).

P815 (H-2^d) cells were incubated for 6 h at 37 °C with the [¹²⁵I]IASA-YIPSAEKI(biotin)I probe and then photoactivated. Analysis of the total cell lysate by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) revealed a major labelled component with an apparent relative molecular mass of 45,000 (*M*, 45K) and a minor one of 68K (Fig. 2, lane 1). The latter could be eliminated by preclearing with a BSA-specific monoclonal antibody (data not shown). The 45K labelled material in the lysate could be recovered nearly quantitatively by immunoprecipitation with a K^d-specific antibody, but not with an antibody specific for D^d or L^d (lanes 2–4). Similar results were obtained with the [¹²⁵I]IASA-YIPSAEKI probe (not shown). The 45K species was also detected by direct SDS-PAGE analysis in lysates from L cells (H-2^k) transfected with the K^d gene, but not in controls from transfectants expressing L^d or D^d (lanes 5–7). Furthermore, labelling was not significant in lysates of EL-4 or CH-27 cells expressing the class I molecules D^b and K^b, or D^k and K^k, respectively (lanes 8 and 9).

Photoaffinity labelling of K^d molecules on P815 cells with [¹²⁵I]IASA-YIPSAEKI(biotin)I was completely inhibited by the unmodified CS P.b. peptide and by other peptides presented by K^d, but was not inhibited by D^b or L^d restricted peptides (Fig. 3). These results are in agreement with those obtained in the functional competition assay (Fig. 1) and with the MHC restriction patterns described for these peptides^{14–19}.

The specificity of MHC-peptide interactions has been investigated in a solid-phase binding assay using solubilized human

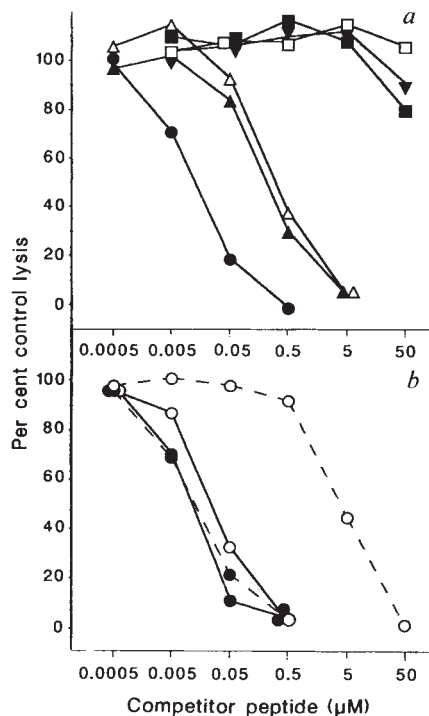


FIG. 1 Comparison of various antigenic peptides and photoreactive peptide conjugates as competitors in a functional competition assay. Peptide CS P.b. 253-260 (●—), P198¹⁴⁻²⁴ (▲—), AYP₅TLA (△—), P91A¹²⁻²⁴ (■—), LCMV NP 118-132 (□—), Ad5 E1a 234-243 (▼—) and photoreactive conjugates IASA-YIPSAEKI (○—), IASA-YIPSAEK(biotin)I (●—), and IASA-YIPSAEKI (○—) were incubated with ⁵¹Cr-labelled P815 target cells at the indicated (final) concentrations, together with the antigenic peptide CW3 170-182 (0.1 μM) and CTL-CW3/1.1 at a 5:1 CTL to target ratio.

METHODS. Peptides were synthesized on a RAMPs peptide synthesizer (Du Pont) using the (9-fluorenylmethoxycarbonyl) (Fmoc) strategy and purified on C-18 reverse-phase HPLC (Waters, Milford, MA). YIPSAEK(biotin)I was prepared by reacting Fmoc-YIPSAEKI with *N*-hydroxysuccinimide biotin (biotin-ONSu) in dimethylsulphoxide in the presence of 1-hydroxybenzotriazole. The reaction mixture was treated with piperidine and the resulting YIPSAEK(biotin)I was purified by C-18 HPLC. The photoreactive derivatives were prepared essentially as described^{29,30}. Freshly iodinated IASA-ONSu was reacted with YIPSAEK(biotin)I and the resulting IASA-YIPSAEK(biotin)I was purified by C-18 HPLC. To prepare the IASA-YIPSAEKI derivative, IASA-ONSu was reacted with the fully protected (except for the N terminus) CS P.b. peptide on the solid-phase resin and the product was then cleaved from the resin and purified by C-18 HPLC. All procedures involving photoreactive reagents were performed under dimmed light. Details of the competition assay are described in ref. 13. The ⁵¹Cr-release assay was terminated after a 4-h incubation at 37 °C. The per cent control lysis was calculated as 100 × (% lysis with competitor minus background lysis) divided by (% lysis without competitor minus background lysis). Background lysis was less than 10%, and lysis without competitor was 54% and 63%, respectively, in a and b.

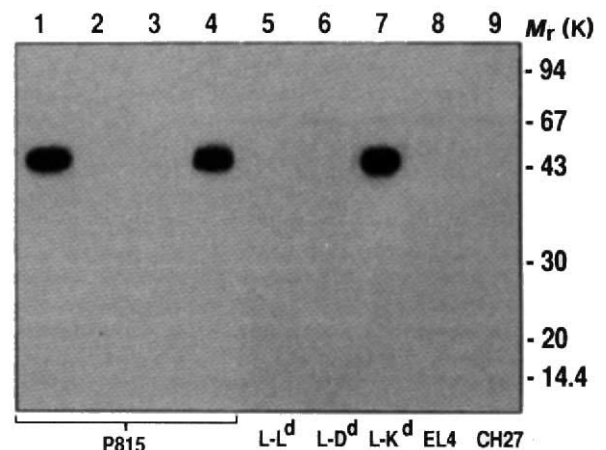
class I (HLA) molecules and peptides immobilized onto plastic⁸⁻¹¹. Although this showed that class I molecules can indeed bind to antigenic peptides, the specificity of binding for many peptides was degenerate. Our experiments, which are based on photoaffinity labelling of class I molecules on living cells, establish that binding of an antigenic peptide to class I molecules can be highly restricted (Figs 1-3). These findings are in agreement with those demonstrating that endogenously processed peptides display a high degree of MHC allel-specificity²⁰⁻²².

As shown in Fig. 4, the photoaffinity labelling of K^d on living P815 cells at 37 °C increased steadily with the incubation time, whereas only weak labelling was observed at 4 °C. A similar reduction in labelling was observed at 37 °C in the presence of sodium azide and 2-deoxyglucose, the protein synthesis inhibitor cycloheximide, or brefeldin A, an inhibitor of protein transport from the endoplasmic reticulum²³ (Fig. 4c, d and e). Reduction

in labelling was not due to a depletion of K^d molecules on the cell surface because flow cytometric analysis confirmed that cell-surface K^d expression was not substantially decreased for at least 6 h under the different incubation conditions (data not shown). Thus the accumulated labelling of K^d molecules on living P815 cells at 37 °C cannot simply be accounted for by direct binding of the photoprobe to the K^d cell-surface molecules available at the start of the incubation. As there was weak labelling in P815 cell lysates at 37 °C (Fig. 4f) and at 4 °C (data not shown), it is unlikely that this binding involves only intracellular K^d molecules available at the beginning of the incubation. The accumulated binding of exogenous peptides to cell-associated class I molecules therefore seems to depend on metabolically active processes. These results are compatible with the interpretation that the peptide photoprobe binds predominantly to newly synthesized class I molecules. As already suggested²⁴⁻²⁷, exogenous peptides might bind to 'nascent' class I heavy chains

FIG. 2 The IASA-YIPSAEK(biotin)I conjugate selectively labels the K^d molecule. P815 cells (H-2^d, lanes 1-4), L cells (H-2^k) transfected with L^d (lane 5), D^d (lane 6) or K^d (lane 7) genes, EL-4 cells (H-2^b, lane 8) and CH-27 cells (H-2^k, lane 9) were incubated with [¹²⁵I]IASA-YIPSAEK(biotin)I and then irradiated with ultraviolet light. Cell lysates were analysed directly by SDS-PAGE (lanes 1, 5-9). In addition, lysates from labelled P815 cells were immunoprecipitated with monoclonal antibody specific for L^d (lane 2), D^d (lane 3) and K^d (lane 4) and then analysed by SDS-PAGE. Migration of size markers is shown on the right.

METHODS. Cells were washed twice in phosphate-buffered saline (PBS) and resuspended in DMEM supplemented with HEPES (20 mM), gentamycin (40 μg ml⁻¹) and FCS (0.5%) at 2 × 10⁶ cell per ml. Cells (2 × 10⁶) were incubated with [¹²⁵I]IASA-YIPSAEK(biotin)I (1 × 10⁷ c.p.m. in 10 μl PBS) in 6-well plates (Costar) at 37 °C for 6 h and then irradiated with ultraviolet light at 4 °C for 10 min (15 watt lamp with an emission maximum at 312 nm; Bioblock Scientific, France). [¹²⁵I]IASA-YIPSAEK(biotin)I was prepared using [¹²⁵I]IASA-ONSu which was prepared as described^{29,30}, had a specific activity of 1,700-2,300 Ci per mmol and was used at a concentration of ~3 × 10⁻⁹ M. Labelled cells were washed twice in cold DMEM containing 5% FCS and once in PBS, then boiled for 5 min in 40 μl sample-reducing buffer. Alternatively, cells on ice were solubilized with NP-40 (0.7%) in the presence of iodoacetamide (10 mM), PMSF (1 mM) and leupeptin (10 μg ml⁻¹), and immunoprecipitated^{29,30}. The anti-K^d monoclonal antibody was SF 1.1.1.1, the anti-D^d was 34-4-20S and the anti-L^d was 28-14-8S. The hybridomas producing these monoclonal antibodies were from American type culture



collection (Rockville, MD). SDS-PAGE was on linear 10% polyacrylamide gels under reducing conditions. Cell-surface expression of L^d, D^d and K^d on the L-cell transfectants^{31,32} was verified by flow cytometric analysis to be at least as high as that on P815 (H-2^d) cells (not shown). For P815 cells, about 300-500 K^d molecules per cell became photolabelled under these conditions.

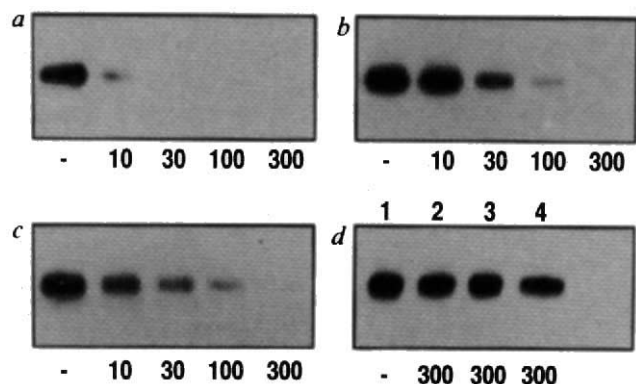


FIG. 3 Competition of the photoaffinity labelling of K^d on P815 cells by different antigenic peptides. P815 cells were incubated as described for Fig. 2 with [125 I] IASA-YIPSAEK(biotin)I in the absence or presence of the indicated molar excess of peptides CS P.b. 253-260 (a), P198.14-24 (b), AYP5TLA (c), LCMV NP 118-132 (d, 2), Ad5 E1a 234-243 (d, 3) and P91A.12-24 (d, 4). After irradiation with ultraviolet light, the material immunoprecipitated by the anti- K^d SF1.1.1 was analysed by SDS-PAGE as described for Fig. 2.

intracellularly, or to newly synthesized 'empty' class I molecules as they emerge at the cell surface.

Photoactivation of the IASA group involves the formation of a short-lived and highly reactive intermediate, phenylnitrene, which can then form covalent bonds with residues on the target protein²⁸. The inclusion of a biotin moiety in the photoprobe, as in the IASA-YIPSAEK(biotin)I compound, should facilitate investigations of the subcellular localization and mechanism of peptide-MHC interactions. Furthermore, studies with class II molecules^{29,30} have demonstrated that photoaffinity labelling can be useful for identifying antigen contact sites on MHC

molecules, and it may now be possible to apply the same approach to MHC class I molecules. □

Received 12 November 1990; accepted 11 March 1991.

1. Townsend, A. & Bodmer, H. *Rev. Immun.* **7**, 601-624 (1989).
2. Schwartz, R. H. *Adv. Immun.* **38**, 31-199 (1987).
3. Buus, S., Sette, A., Colon, S. M., Miles, C. & Grey, H. M. *Science* **235**, 1353-1362 (1987).
4. Babbitt, B. P., Allen, P. M., Matsueda, G., Haber, E. & Unanue, E. R. *Nature* **317**, 359-361 (1985).
5. Buus, S. *et al. Proc. natn. Acad. Sci. U.S.A.* **83**, 3968-3971 (1986).
6. Schaeffer, E. B. *et al. Proc. natn. Acad. Sci. U.S.A.* **86**, 4649-4653 (1989).
7. Chen, B. P. & Parham, P. *Nature* **337**, 743-745 (1989).
8. Bouilliot, M. *et al. Nature* **339**, 473-475 (1989).
9. Frelinger, J. A., Gotch, F. M., Zweernink, H., Wain, E. & McMichael, A. J. *J. exp. Med.* **172**, 827-834 (1990).
10. Choppin, J. *et al. J. exp. Med.* **172**, 889-899 (1990).
11. Chen, B. P., Rothbard, J. & Parham, P. *J. exp. Med.* **172**, 931-936 (1990).
12. Pala, P. *et al. J. Immun.* **141**, 2289-2294 (1988).
13. Maryanski, J. L., Pala, P., Cerottini, J.-C. & Corradin, G. *J. exp. Med.* **167**, 1391-1405 (1988).
14. Romero, P. *et al. Nature* **341**, 323-326 (1989).
15. Sibille, C. *et al. J. exp. Med.* **172**, 35-45 (1990).
16. Maryanski, J. L., Verdini, A. S., Weber, P. C., Salemme, F. R. & Corradin, G. *Cell* **60**, 63-72 (1990).
17. Lurquin, C. *et al. Cell* **58**, 293-303 (1989).
18. Schulz, M. *et al. Eur. J. Immun.* **19**, 1657-1667 (1989).
19. Kast, W. M. *et al. Cell* **59**, 603-614 (1989).
20. Van Bleek, G. M. & Nathanson, S. G. *Nature* **348**, 213-216 (1990).
21. Falk, K., Rötzschke, O. & Rammensee, H.-G. *Nature* **348**, 248-251 (1990).
22. Rötzschke, O. *et al. Nature* **348**, 252-254 (1990).
23. Misumi, Y. *et al. J. Biol. Chem.* **261**, 11398-11403 (1986).
24. Townsend, A. *et al. Nature* **340**, 443-448 (1989).
25. Ljunggren, H.-G. *et al. Nature* **346**, 476-480 (1990).
26. Schumacher, T. N. M. *et al. Cell* **62**, 563-567 (1990).
27. Townsend, A. *et al. Cell* **62**, 285-295 (1990).
28. Bayley, H. & Staros, J. V. in *Azides and Nitrenes—Reactivity and Utility* (ed. Scriven, E. F. V.) 434-487 (Academic, London, 1984).
29. Luescher, I. F., Allen, P. M. & Unanue, E. R. *Proc. natn. Acad. Sci. U.S.A.* **85**, 871-874 (1988).
30. Luescher, I. F., Crimmins, D. L., Schwartz, B. D. & Unanue, E. R. *J. Biol. Chem.* **265**, 11177-11184 (1990).
31. Maryanski, J. L., Accolla, R. S. & Jordan, B. J. *Immun.* **136**, 4340-4347 (1986).
32. Maryanski, J. L., Abastado, J.-P. & Kourilsky, P. *Nature* **330**, 660-662 (1987).

ACKNOWLEDGEMENTS. We thank G. Corradin for several synthetic peptides, B. Malissen for the L- D^d transfectant, K. Mühlethaler for technical assistance and A. Zoppi for preparing the manuscript. P.R. was supported, in part, by the Roche Research Foundation.

Peptide binding to empty HLA-B27 molecules of viable human cells

Richard J. Benjamin, J. Alejandro Madrigal & Peter Parham

Departments of Cell Biology and Microbiology & Immunology, Stanford University, Stanford, California 94305-5400, USA

INTRACELLULAR binding of antigenic peptides by polymorphic class I major histocompatibility complex molecules creates the ligands recognized by receptors of CD8⁺ T cells¹. Previously described *in vitro* assays of peptide binding to class I molecules have been limited by either the low proportion of accessible binding sites or the lack of allelic specificity²⁻⁶. Here we describe a system in which the human class I molecule HLA-B27 binds considerable amounts of an influenza peptide with precise allelic discrimination. Binding requires viable cells, is stimulated by γ -interferon and is inhibited by brefeldin A. Our results are consistent with the presence of fairly stable 'empty' HLA-B27 molecules at the cell surface. By contrast, analysis of the binding of a second influenza peptide indicates that empty HLA-Aw68 molecules are relatively short-lived. We speculate that this property could underlie its association with autoimmune disease.

HLA-B27 presents a synthetic peptide (NP1) derived from residues 383-394 of the influenza nucleoprotein to T cells⁷. We cultured radioiodinated NP1 with HLA-B27-transfected C1R cells (an HLA-A, B-negative human B-cell line⁸) which were concomitantly radiolabelled with [35 S]methionine. Measurable [125 I]-labelled peptide could be immunoprecipitated with class I molecules (Fig. 1a); all three components (HLA-B27 heavy

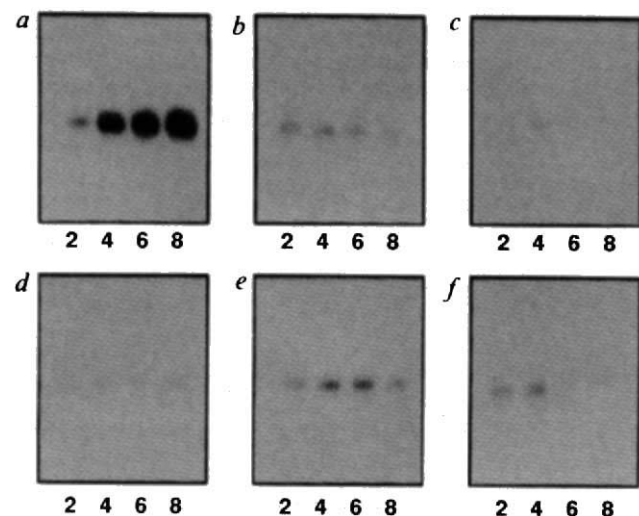


FIG. 4 Photoaffinity labelling of K^d on P815 cells in the presence or absence of different inhibitors. Cells were incubated with [125 I]IASA-YIPSAEK(biotin)I for the indicated periods of time at 4 °C (b) or at 37 °C (a, c-f) in the absence (a) or presence of sodium azide (0.02%) and 2-deoxyglucose (10 mM) (c), cycloheximide (10 μ g ml⁻¹) (d), or brefeldin A (1 μ g ml⁻¹) (e). Alternatively, the cells were first solubilized with MEGA 8 (30 mM) and MEGA 9 (30 mM) detergents plus a mixture of enzyme inhibitors (see methods for Fig. 2) immediately before incubation with the photoprobe (f). Following ultraviolet irradiation, the material immunoprecipitated with the anti- K^d antibody SF1.1.1.1 was analysed by SDS-PAGE as described for Fig. 2.

FIG. 1 *a, b*, NP1 peptide reproducibly coprecipitates with HLA-B*2705. 125 I-NP1 was cultured with either untransfected C1R cells (solid), B27-C1R transfectants (open), or B7-C1R (stipple) cells, and the HLA-class I molecules immunoprecipitated and counted for radioactivity. *b*, 12.5% SDS-PAGE of individual precipitates. *c*, Reverse-phase HPLC analysis of 125 I-NP1 (dotted line) and peptide eluted from HLA-B*2705 immunoprecipitates (solid line). *d*, Peptide binding can be inhibited by an excess of HLA-B*2705 restricted peptides. 125 I-NP1 ($2\mu\text{g ml}^{-1}$) was incubated with B27-C1R or C1R cells (cross) plus cold peptides. These inhibitors and their amino-acid sequences in the one-letter amino-acid code were NP1, SRYWAIKTRSGG (open squares); HIVgag (265-279) peptide, KRWIIIGLNKIVRMVYC (solid squares); FluNP (89-101), PKKTGGPIYKRVD (open diamonds); or FluNP (335-349), SAAFDLRVLSFIRGY (solid diamonds). The carboxy-terminal residues of HIVgag peptide and FluNP (335-349) were not a part of the viral sequences. METHODS. Iodination was by the chloramine T method, the product being separated on a Sep-Pak C18 cartridge. Typically, 2-8% of molecules were labelled. HLA class-I expression was determined on the day of each assay by immunofluorescence. Cells were preincubated overnight with 500 units per ml γ -interferon. 125 I-NP1 peptide was added to 2.5×10^6 cells in 0.5 ml methionine-free RPMI plus 1% BSA, 500 units per ml γ -interferon and $60\mu\text{Ci ml}^{-1}$ ^{35}S -methionine, and incubated for 6 h at 37°C . Cells were washed and lysed, and the class I molecules immunoprecipitated and washed as described¹⁹. Lysates were precleared twice for 30 min with 125 μl washed Staph A cells plus 5 μl normal rabbit serum. Immunoprecipitation was effected with 10 μg purified W6/32 (ref. 29) plus 50 μl Staph A cells and incubation for 30 min on ice. Precipitated radioactivity was measured directly with a LKB 1260 Multigamma II Gamma Counter. Peptide was eluted from immunoprecipitates in 70% acetonitrile, 0.3% trifluoroacetic acid (TFA), then dried, redissolved in 0.1% TFA and loaded onto an Altex Ultrasphere ODS C18 column, from which it was eluted with a 0-80% acetonitrile gradient in 0.1% TFA.

chain, β 2-microglobulin (β 2m) and NP1) were resolved by SDS-polyacrylamide gel electrophoresis (Fig. 1*b*). Reverse-phase HPLC of the bound radioiodinated peptide revealed a single dominant peak that was distinct from the main components of the incubated 125 I-labelled peptide (Fig. 1*c*), indicating that either HLA-B27 is selecting a minor component of the added peptide or that proteolytic processing of the peptide is occurring⁹⁻¹¹.

This binding had peptide and HLA specificity corresponding to the patterns of cytotoxic T lymphocyte recognition⁷. It was inhibited by non-radioactive preparations of NP1 and an HLA-B27-presented human immunodeficiency virus gag peptide¹², but not by other peptides (Fig. 1*d*). Similarly, 125 I-NP1 only bound to the common subtype of HLA-B27 (B*2705) and not to seven other HLA-A, B molecules (Fig. 2*a, b*). Binding was not due to nonspecific 'stickiness' of HLA-B27 because a second peptide (NP2) derived from residues 335-349 of the influenza nucleoprotein¹³ bound to HLA-Aw68 (A*6801) but not to HLA-B27 (Fig. 2*c*). Neither was binding peculiar to transfectants of the mutant C1R cell line as non-mutant B-cell lines gave similar results (Fig. 3*a, b*). That binding is not simply a function of the cell-surface abundance of HLA-B27 molecules was shown by the increased binding to C1R transfectants in comparison to normal B-cell lines expressing equivalent or greater amounts of cell-surface HLA-B27, and also by the 50-100% increase in peptide binding produced by γ -interferon compared with its minimal effect ($\pm 10\%$) on cell-surface expression (Fig. 3*c*). No change in peptide binding was caused by the presence of serum or by the addition of purified human β 2m (data not shown), thus β 2m exchange does not seem to be a significant factor in the peptide-binding mechanism¹⁴⁻¹⁶.

Inhibition by cycloheximide and brefeldin A (refs 17, 18) showed that biosynthesis and exocytic transport, respectively, of class I polypeptides are important in this assay. Preincubation and inclusion of either drug in the assay reduced binding of NP1 to HLA-B27 by $\sim 50\%$, and of NP2 to HLA-Aw68 by 80-90% (Fig. 4). Binding of NP1 to HLA-B27 is not due to reactions taking place after cell culture in the detergent cell lysate, as it was not blocked by addition of excess 'cold' peptide after cell culture and before lysis (Fig. 4*b,ii*). Neither was binding reconstituted by addition of 125 I-NP1, nor of C1R cells loaded

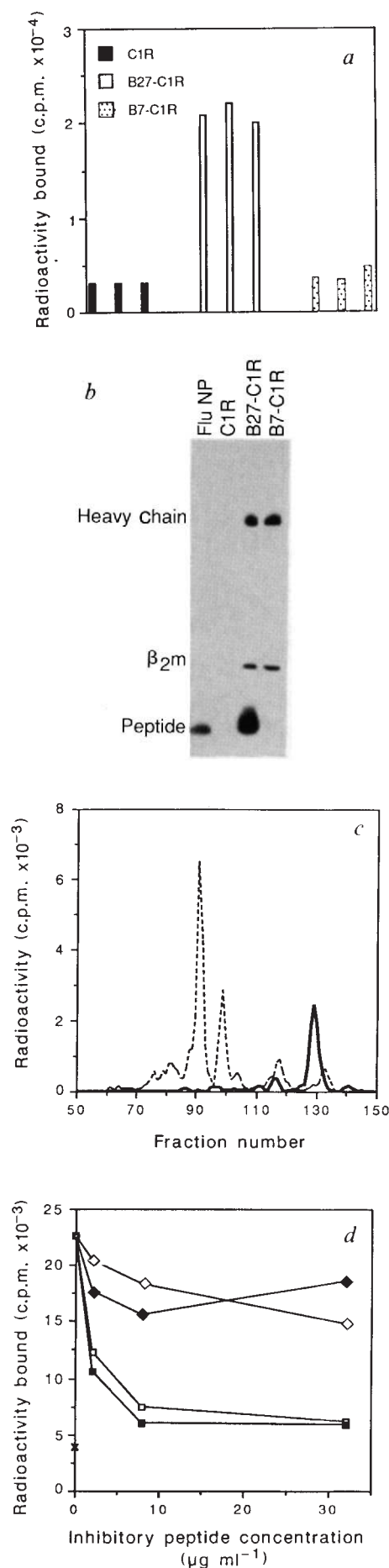


FIG. 2 Allelic specificity of peptide binding. I^{125} -NP1 peptide binding to C1R cells expressing *a*, different HLA-A,B alleles and *b*, different HLA-B27 subtypes³⁰; *c* I^{125} -NP2 binding to C1R cells expressing different class I alleles. METHODS. B*701-C1R, A*201-C1R and A*6801-C1R were stable transfectants. HLA-B27 subtypes were produced by site-directed mutagenesis of the HLA-B*2705 genomic sequence by the method of Kunkel³¹ using the Biorad (Richmond, CA) Muta-gene *in vitro* mutagenesis kit. The following mutations were made: HLA-B*2701 (74D→Y; 77D→N; 81L→A); HLA-B*2702 (77D→N; 80T→A; 81L→A); HLA-B*2703 (59Y→H); HLA-B*2706 (77D→S; 114H→D; 116D→Y; 152V→E) and were based on the sequences described in ref. 30. Mutant and wild-type genes were subcloned into the vector pHEBO and electroporated into C1R cells as described³². HLA class-I expression was determined by immunofluorescence: all transfectants gave comparable levels on the day of assay. Isoelectric focusing gel electrophoresis confirmed the mutant gene products had isoelectric points identical to the naturally occurring HLA-B27 subtypes.

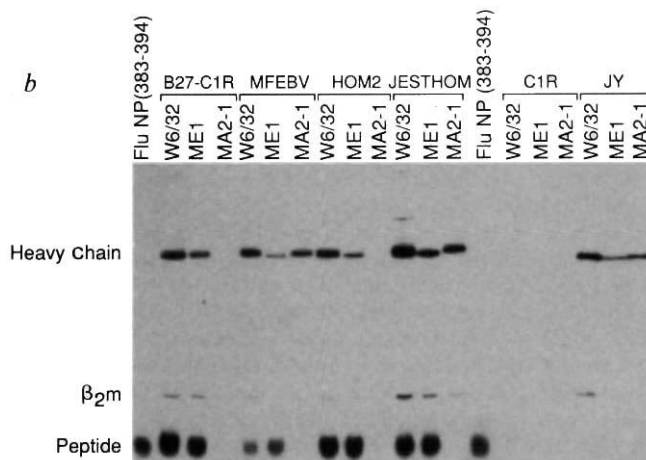
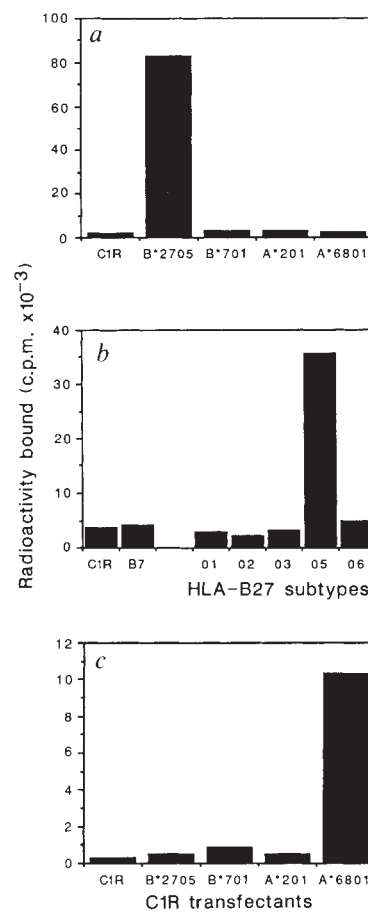
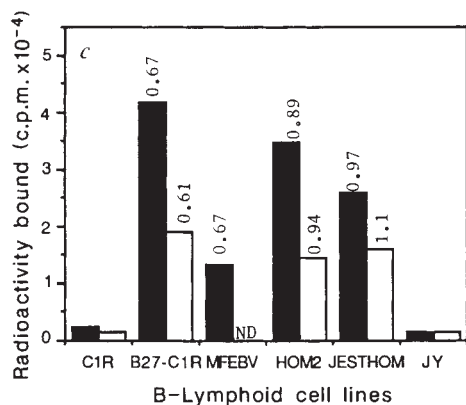
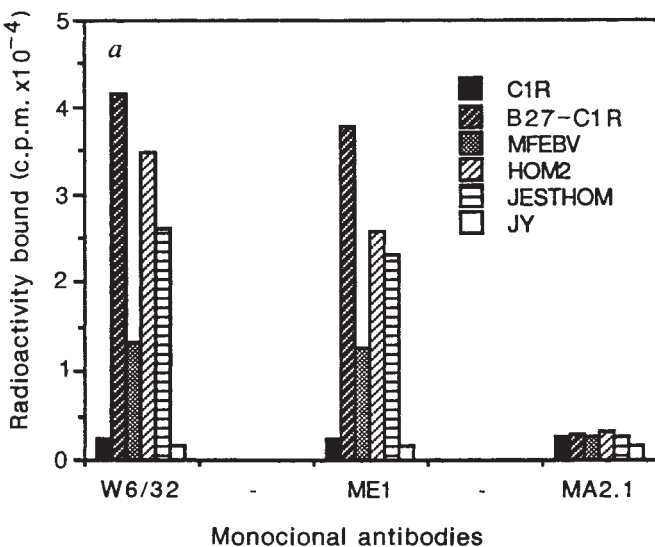
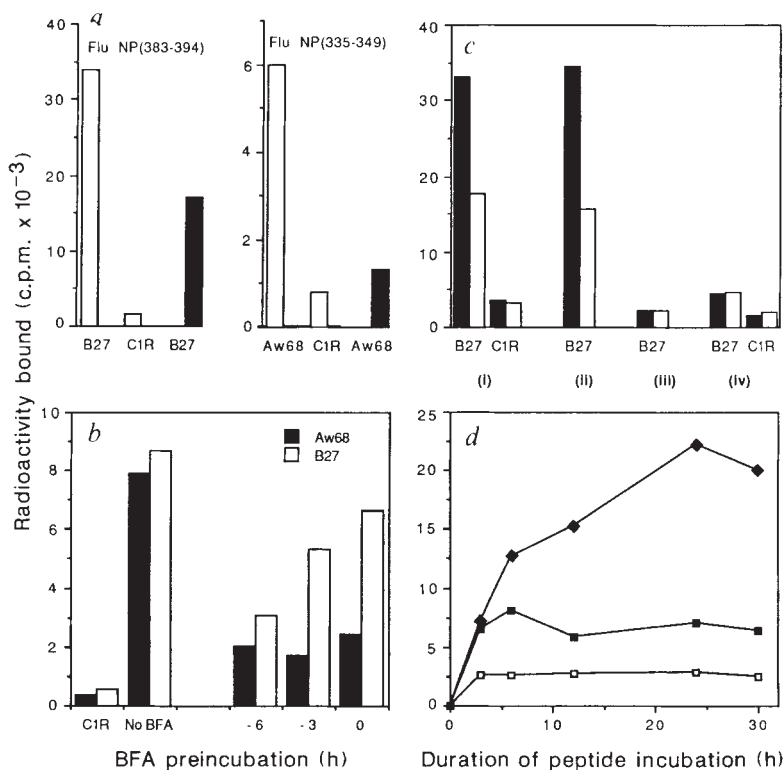


FIG. 3 *a*, Peptide binding to Epstein-Barr virus (EBV)-transformed human B cell lines. The binding of I^{125} -NP1 peptide to HLA-B*2705 positive (MF, HOM2 and JESTHOM) and HLA-B*2705 negative (JY) B-cell lines was assessed. B27-C1R and C1R cells were included as positive and negative controls. *b*, Reducing 12.5% SDS-PAGE of immunoprecipitates from EBV-transformed B-cell lines. *c*, Effect of γ -interferon on I^{125} -NP1 binding to B-cell lines. Cells were incubated either in the presence of 500 units per ml γ -interferon (solid bars) or in its absence (open bars) from both the overnight preincubation and the peptide incubation. Numbers represent the HLA-B*2705 cell-surface expression as determined by quantitative radioimmunoassay ($\times 10^6$) (ref. 33).

METHODS. HLA types were: C1R (–, –, Cw4); B27-C1R (–, B*2705, Cw4); MF (A1, A2, B*2705, Bw54, Cw1); HOM2 (A3, B*2705, Cw1); JESTHOM (A2, B*2705, Cw1) and JY (A2, B7). HLA class I was immunoprecipitated with either W6/32 (monomorphic), ME1 (specific for HLA-B27, B7, Bw42 and Bw22) (ref. 34) or MA2.1 (specific for HLA-A2 and B17) (ref. 35).



with ^{125}I -NP1, to mock-cultured B27-C1R cells before lysis (Fig. 4c, iii and iv).

In the presence of brefeldin A, maximal binding of ^{125}I -NP1 to HLA-B27 was reached within 6 hours. Without brefeldin A, a similar initial component was observed, followed by a slower component that attained a higher maximum after 24 hours. This level corresponds to 2.4×10^5 peptide-loaded HLA-B27 molecules per cell, about 36% of the total cell-surface HLA-B27 molecules detected by radioimmune assay. These results indicate that exocytosis of newly synthesized HLA-B27 is the rate-limiting step in peptide binding.

Our results are consistent with the interpretation that the observed binding is to functionally empty class I HLA molecules¹⁹⁻²⁴. We have not demonstrated that binding is at the cell surface, and an alternative intracellular site for peptide loading is not ruled out. The binding detected in the presence of brefeldin A may reflect a relative stability of empty HLA-B27 molecules. Increasing the period of preincubation with brefeldin A diminishes the magnitude of this binding, but it was still detectable even after a 6-hour preincubation. Peptide binding to HLA-Aw68 was more sensitive to brefeldin A and cycloheximide, indicating empty HLA-B27 molecules are longer lived than empty HLA-Aw68 molecules. Comparison with previous reports suggests that empty HLA-B27 molecules are also more stable than empty H-2 molecules¹⁹⁻²⁴. A further distinction is that culture at reduced temperature (26 °C) does not increase peptide binding to empty HLA-B27 molecules. By contrast, overnight preincubation with γ -interferon increases peptide binding even when brefeldin A is added later (data not shown). As previously proposed²⁵, this suggests that γ -interferon facilitates assembly and exocytosis of class I molecules—in this case, empty class I molecules.

With this 'peptide feeding' assay we find that a significant proportion of cell-associated HLA-B27 molecules bind peptide with allelic specificity. Peptides seem to bind to empty HLA-B27 molecules that are relatively stable under physiological conditions. The possibility that certain class I molecules present extracellular peptides to T cells *in vivo* should therefore be

FIG. 4 a, Effect of cycloheximide (solid bars) on ^{125}I -NP1 binding to HLA-B*2705 and ^{125}I -NP2 binding to HLA-A*6801. b, Effect of brefeldin A (BFA), given for varying preincubation periods, on ^{125}I -NP1 binding to HLA-B*2705 (open columns) and ^{125}I -NP2 binding to HLA-A*6801 (solid columns). c, ^{125}I -NP1 binding, in the presence (open bars) or absence (solid bars) of brefeldin A: (i) control incubation with $4 \mu\text{g ml}^{-1}$ ^{125}I -NP1; (ii) $10 \mu\text{g}$ non-radioactive NP1 added to washed cells just before lysis; (iii) B27-C1R cells mock-treated without peptide for 6 h at 37 °C, washed, and then mixed with washed ^{125}I -NP1 laden C1R cells just before lysis; (iv) B27-C1R or C1R cells mock-labelled without peptide, washed and $100 \text{ ng } ^{125}\text{I}$ -NP1 added just before cell lysis. d, Binding of ^{125}I -NP1 as a function of time. Peptide binding at $20 \mu\text{g ml}^{-1}$ to untreated B27-C1R transfectants (solid diamonds), brefeldin A-treated B27-C1R transfectants (solid squares) and to C1R cells (open squares).

METHODS. Cycloheximide was added to $20 \mu\text{M}$ one hour before ^{125}I -NP1. SDS-PAGE confirmed the complete inhibition of ^{35}S -methionine incorporation into class I molecules. Brefeldin A was resuspended at 10 mg ml^{-1} in methanol and stored at -70°C . A titration determined that maximum inhibition of peptide binding, and also of sialation of HLA class I (as determined by isoelectric focusing gel electrophoresis) occurred at concentrations of $1 \mu\text{g ml}^{-1}$. Cells were pretreated for 2 h with $10 \mu\text{g ml}^{-1}$ brefeldin A before addition of ^{125}I -NP1 and the standard 6-h culture in the experiments shown in Fig. 4c and d. In c, iv, $100 \text{ ng } ^{125}\text{I}$ -NP1 was used as this represented the maximum percentage (5%) of peptide generally carried over in cell lysates after washing. The percentage surface class I labelled was determined from: Average number of peptide molecules per cell = (total no. peptides bound)/(total c.p.m. bound) $\times$ (moles peptide $\times$ Avogadro's no.)/(total c.p.m. ^{125}I -NP1).

considered. For HLA-B27, this property might underlie the strong association of this allele with ankylosing spondylitis and other seronegative spondylarthropathies (refs 26 and 27; reviewed in ref. 28).

Received 4 December 1990; accepted 11 March 1991.

1. Townsend, A. & Bodmer, H. A. *Rev. Immun.* **7**, 601-624 (1989).
2. Chen, B. P. & Parham, P. *Nature* **337**, 743-745 (1989).
3. Bouillot, M. *et al. Nature* **339**, 473-475 (1989).
4. Frelinger, J. A., Gotch, F. M., Zweerink, H., Wain, E. & McMichael, A. J. *J. exp. Med.* **172**, 827-834 (1990).
5. Choppin, J. *et al. J. exp. Med.* **172**, 889-899 (1990).
6. Chen, B. P., Rothbard, J. & Parham, P. *J. exp. Med.* **172**, 931-936 (1990).
7. Huet, S. *et al. Int. Immun.* **2**, 311-316 (1990).
8. Storkus, W. J., Howell, D. N., Salter, R. D., Dawson, J. R. & Cresswell, P. *J. Immun.* **138**, 1657-1659 (1987).
9. Van Bleek, G. M. & Nathanson, S. G. *Nature* **348**, 213-216 (1990).
10. Falk, K., Rötzschke, O. & Rammensee, H.-G. *Nature* **348**, 248-251 (1990).
11. Rötzschke, O. *et al. Nature* **348**, 252-254 (1990).
12. Nixon, D. F. *et al. Nature* **336**, 484-487 (1988).
13. McMichael, A. J., Gotch, F. M. & Rothbard, J. *J. exp. Med.* **164**, 1397-1406 (1986).
14. Rock, K. L., Rothstein, L. E., Gamble, S. R. & Benacerraf, B. *Proc. natn. Acad. Sci. U.S.A.* **87**, 7517-7521 (1990).
15. Vitiello, A., Potter, T. A. & Sherman, L. A. *Science* **250**, 1423-1426 (1990).
16. Kozlowski, S. *et al. Nature* **349**, 74-77 (1991).
17. Nuchtern, J. G., Bonifacio, J. S., Biddison, W. E. & Klausner, R. D. *Nature* **339**, 223-226 (1989).
18. Yewdell, J. W. & Bennink, J. R. *Science* **244**, 1072-1075 (1989).
19. Townsend, A. *et al. Nature* **340**, 443-448 (1989).
20. Cerundolo, V. *et al. Nature* **345**, 449-452 (1990).
21. Lie, W.-R. *et al. Nature* **344**, 439-441 (1990).
22. Ljunggren, H.-G. *et al. Nature* **346**, 476-480 (1990).
23. Townsend, A. *et al. Cell* **62**, 285-295 (1990).
24. Schumacher, T. N. M. *et al. Cell* **62**, 563-567 (1990).
25. Klar, D. & Hämmerling, G. J. *EMBO J.* **8**, 475-481 (1989).
26. Brewerton, D. A. *et al. Lancet* **i**, 904-907 (1973).
27. Schlosstein, L., Terasaki, P. I., Bluestone, R. & Pearson, C. M. *New Engl. J. Med.* **288**, 704-706 (1973).
28. Benjamin, R. J. & Parham, P. *Immun. Today* **11**, 137-142 (1990).
29. Barnstable, C. J. *et al. Cell* **14**, 9-20 (1978).
30. Lopez de Castro, J. A. *Immun. Today* **10**, 239-246 (1989).
31. Kunkel, T. A. *Proc. natn. Acad. Sci. U.S.A.* **82**, 488-492 (1985).
32. Salter, R. D., Clayberger, C., Lomen, C. E., Krensky, A. M. & Parham, P. *J. exp. Med.* **166**, 283-288 (1987).
33. Ways, J. P. & Parham, P. *Biochem. J.* **216**, 423-432 (1983).
34. Ellis, S. A., Taylor, C. & McMichael, A. *Hum. Immun.* **5**, 49-59 (1982).
35. McMichael, A. J., Parham, P., Rust, N. & Brodsky, F. *Hum. Immun.* **1**, 121-129 (1980).

ACKNOWLEDGEMENTS. We thank R. Busch and J. Rothbard for discussion and amino-acid analyses. Supported by grants from the USPHS. R.J.B. is a visiting fellow funded by the Lucille P. Markey Trust. J.A.M. is a fellow of the Leukemia Society of America.

Acquisition of the human adeno-associated virus type-2 *rep* gene by human herpesvirus type-6

Brian J. Thomson, Stacey Efsthathiou*
& Robert W. Honess

Division of Virology, National Institute for Medical Research,
The Ridgeway, Mill Hill, London NW7 1AA, UK

* Division of Virology, Department of Pathology, University of Cambridge,
Tennis Court Road, Cambridge CB2 1QP, UK

HUMAN herpesvirus type-6 (HHV-6) is a recently isolated herpesvirus which is highly prevalent in adult populations around the world¹⁻³. HHV-6 was first isolated from the peripheral blood of six individuals with lymphoproliferative disorders, two of whom were also infected with human immunodeficiency virus¹. HHV-6, in common with other herpesviruses, transactivates the HIV long terminal repeat linked to reporter genes^{4,5} and has in addition been shown to accelerate HIV gene expression and CD4 cell death in cultures co-infected with both viruses⁶. The virus is tropic for CD4⁺ lymphocytes^{7,8} and persists in the peripheral blood of most seropositive individuals⁹. We have now identified a gene in HHV-6 encoding a 490-amino-acid polypeptide homologous to the human adeno-associated virus type-2 (AAV-2) *rep* gene¹⁰. This gene has an essential role in AAV-2 DNA replication^{11,12}, can *trans*-regulate homologous and heterologous gene expression¹³⁻¹⁷, and inhibits cellular transformation¹⁸. The acquisition of *rep* by HHV-6 could be due to natural transfer of genetic information between DNA viruses of eukaryotes and is likely to have important consequences for the life-cycle of HHV-6 and for the host CD4 cell.

The genomic structure of a Ugandan isolate of HHV-6 (strain U1102) has been determined and consists of a 141-kilobase (kb), largely unique sequence flanked by direct repeat units of roughly 10 kb¹⁹. Sequencing shows that HHV-6 is closely phylogenetically related to human cytomegalovirus²⁰ and that the genetic organization of these two herpesviruses is highly conserved (B.J.T. *et al.*, manuscript submitted). Nucleotide analysis of a 12-kb *Hind*III fragment of HHV-6 U1102, which lies proximal to the right-hand terminal repeat of the viral genome¹⁹ (Fig. 1a), has identified a 490-amino-acid open reading frame (Fig. 1b) homologous to the *rep* 68/78 gene product of the human helper-dependent parvovirus AAV-2 (ref. 10). The product of the HHV-6 *rep* gene homologue (HHV-6 PH) was more closely related to the AAV-2 gene product than were the equivalent non-structural NS1/*rep* proteins encoded by the autonomous parvoviruses (Table 1). An alignment of the predicted translation product of HHV-6 PH with the AAV-2 Rep 68/78 polypeptide is shown in Fig. 2. The proteins share a 24% identity across the entire length of the 490-amino-acid HHV-6 PH reading frame. The close relationship between the two proteins is most clearly illustrated by a DIAGON²¹ plot, compared with a plot showing the relationship of the AAV-2 Rep protein to the equivalent protein encoded by human autonomous parvovirus B19 (Fig. 3a and b). The HHV-6 PH product is more closely related to the AAV-2 protein throughout its length than are the helper-dependent and autonomous parvovirus proteins to each other. All parvovirus Rep/NS1 proteins contain a well conserved 145-amino-acid domain which has significant homology to the T antigens of polyomavirus and simian virus 40 and the E1 proteins of papillomaviruses²². This domain contains a consensus purine nucleotide triphosphate binding site essential to the role of these proteins in DNA replication²³. This consensus is conserved in HHV-6 PH. A multiple alignment of this region of the HHV-6 PH product with AAV-2 and autonomous parvovirus Rep/NS1 proteins is shown in Fig. 3c.

We have no direct evidence for the expression of HHV-6 PH. There is, however, no evidence for the presence of pseudogenes

TABLE 1 Homology between HHV-6 PH, AAV-2 and equivalent proteins

	HHV-6	AAV-2	B19	BPV	MVM
HHV-6	*	619	315	235	210
AAV-2		*	598	417	390
B19			*	399	289
BPV				*	319
MVM					*

FASTA scores obtained by searching the PIR database with the amino-acid sequences of the HHV-6 PH ORF, AAV-2 *rep* 68/78 gene and the equivalent non-structural NS1 proteins of human (B19), bovine (BPV) and murine (MVM) autonomous parvoviruses (refs 28, 30 and 31, respectively). Searches were made using the FASTA program³⁹ using a Ktuple of 1.

in the unique regions of other characterized herpesviruses and the codon usage in this *orf* conforms to that of other recognized HHV-6 genes (ref. 20, and S.E. and B.J.T., unpublished observations). The first ATG codon conforms to the Kozak consensus²⁴; there is a TATA box 28 base pairs (bp) upstream of this methionine codon and a polyadenylation signal sequence AATAAA at position 1,817 (Fig. 1). The HHV-6 PH is therefore highly likely to be expressed.

Herpesvirus genomes are known to contain genes related to ancestral cellular counterparts²⁵⁻²⁷. These genes lack the introns of their cellular homologues, suggesting acquisition by the reverse transcription of mature messenger RNA. The region of the AAV-2 *rep* gene product to which HHV-6 PH is homologous is encoded by unspliced mRNA. The helper-dependent and autonomous parvovirus Rep proteins appear to constitute a unique family of regulatory proteins and no cellular homologue has been identified. The nucleotide sequences of several herpesvirus (summarized in ref. 19) and parvovirus²⁸⁻³¹ genomes do not indicate a common ancestral origin for these two distinct lineages and in particular no evidence of a *rep* gene homologue in other herpesviruses. The HHV-6 PH protein is more closely related to the AAV-2 *rep* 68/78 gene product than are the Rep proteins of the other sequenced parvoviruses, and DIAGON²¹ comparisons of the two proteins indicate that this similarity is not limited to recognized functional domains (Fig. 3a, b). Although the homology may be due to a common ancestral origin or convergence in sequence evolution, the nature of the sequence similarity between the HHV-6 PH protein and the product of the AAV-2 *rep* 68/78 gene indicates that HHV-6 has acquired this gene by direct transfer from AAV-2. The properties of AAV-2 suggest a way in which this could take place. Adeno-associated viruses do not normally replicate in cells unless they have been coinfecting with a member of either the adeno- or herpesvirus families³². In the absence of helper virus, adeno-associated viruses establish latent infection of the host cell as an integrated provirus³³. HHV-6, which is likely to provide such helper function, could therefore have 'captured' the *rep* gene by a mechanism of nonhomologous recombination with AAV-2 replicative intermediates, or integrated provirus.

The AAV-2 *rep* gene encodes at least four overlapping polypeptides produced by the use of two promoters and alternative splicing and identified by their apparent molecular size on acrylamide gels as Rep 78, Rep 68, Rep 52 and Rep 40^{10,11,34}. Rep 78 is encoded by a 4.2-kb unspliced transcript, and the closely related Rep 68 from a singly spliced 3.9-kb transcript. Rep 68 binds to the AAV-2 origin of replication in a sequence-dependent manner and has a well defined ATP-dependent site-specific endonuclease and DNA helicase activity¹¹. The region of homology with HHV-6 includes almost all of the Rep 68 protein. Mutations in the region of the *rep* 68/78 gene homologous to HHV-6 PH show that the gene has other important functions¹²⁻¹⁶. AAV-2 *rep* gene products act as positive and negative regulators of both homologous^{13,14,16} and heterologous¹⁵ promoters, including the HIV long terminal

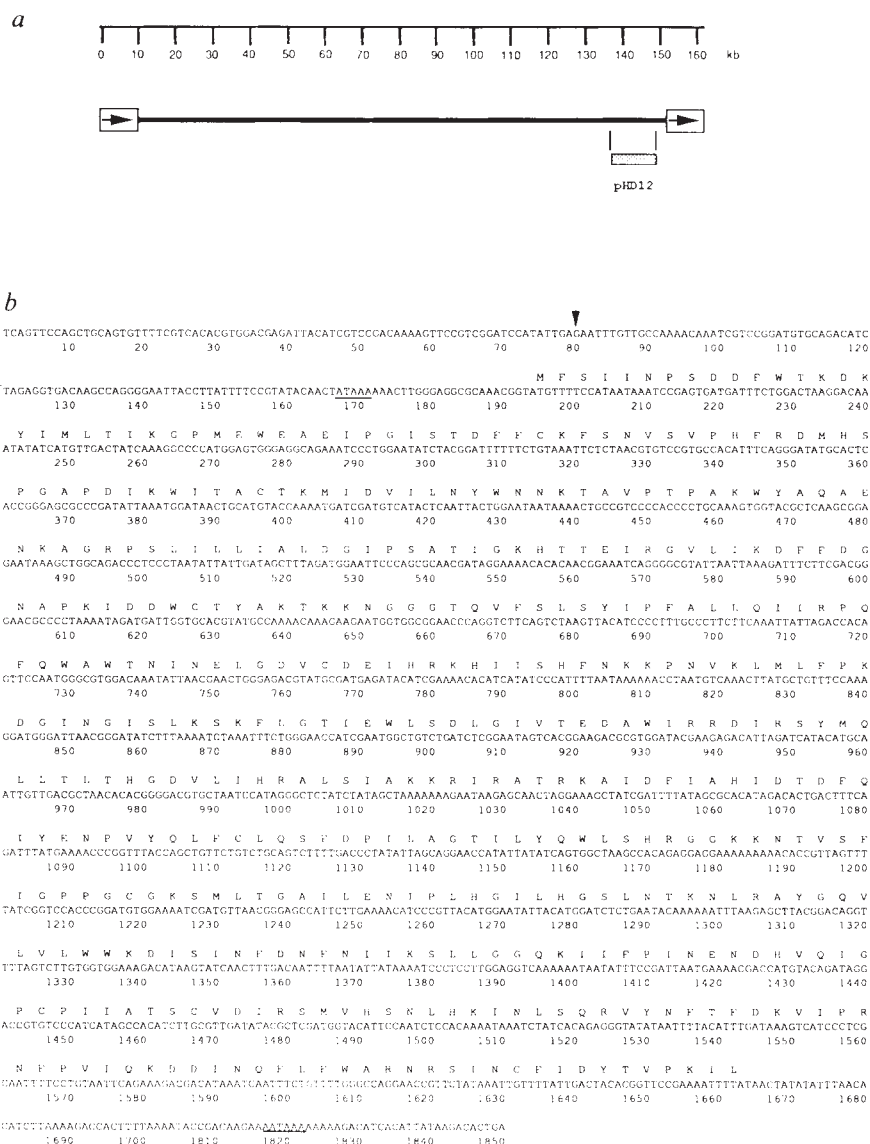


FIG. 2. Amino-acid sequence homology (single-letter amino-acid code) shared between the HHV-6 PH open reading frame and the AAV-2 *rep* 68/78 gene product¹⁰. Boxed residues indicate the identity between the two sequences. The entire 490-amino-acid HHV-6 gene sequence is shown aligned with the first 490 amino acids common to both the Rep 78 and Rep 68 AAV-2 polypeptides. The sequences were aligned manually with the aid of the FASTA program library search alignment output³⁹.

MFSIINPSDDFWTKDKYVIMLT	KPFHEWEAE	PGISTDFCKFSNU---HFRDTHSPG	58
MPGFEVILUKPSDLDGHL	PGISDSE	UNVUAEKEWELPPSDMLNL	47
APDIKUWITACTKXIDU	LYNYHNKTAUPTFKUYA	HNKAGRPSLILL	118
IEQAP-LTUAERLQD	LTUARUSKAPF	ELFFUQFKESVFNHMLUETTGU	105
CKHTTEITAGULIKOFFDGNAP	IDICTYAKTKK	NGGGTQFSLSVIPFALLQIIAPF	177
GRFLSQLAEKLQRIYAGI	EPTLPLNF	FAUTKTNGAGGGGNNKUEQVIMVLLPKTQPEL	165
QUAUTNINELGDUC	DEIHAHHISHF	-----NKKPNULMFLPKDGIN	229
QUAUTNHEQVLSALNLT	AKRLURQHLTHUSQTQ	ENONENHNSDAPULASKTSARYN	225
GTIELLSQIGUTEDU	LIARDISV	MOLLTLTHGDUV	289
ELUGULNKGCTSEKQ	LIQEDDAS	ISFHAASNSASQIKALDNACK	285
HIDTDFQIVENPULFLOS	FDPILAGTILY	QLSHAGGKNTUSF	349
GOQPUEIDSSIAHX	ILENGV	POVARSUFLQATKKFKRNTIULFGATTCTK	345
AILENIPHLGILHGS	LNTKNLRAVGQUL	LLUKDISINFDNFNIKSULGGQK	409
AIAHTUPFGCUNUTN	ENFFNDUCUKN	ILLUEEGKTAKUVESAKILGGS	405
NDHUQIDPPKIPATS	-----	QUHARSMUHNSHKINLSORUYNF	465
KSSAQIDPTEVIL	LSNTNHC	AVLUGNSTTFEHOPLQDHF	465
INQFLPHARASINCF	IDVTPYTL	490	
UKDFRAUKDHUUE	UEHEFVYKGG	490	

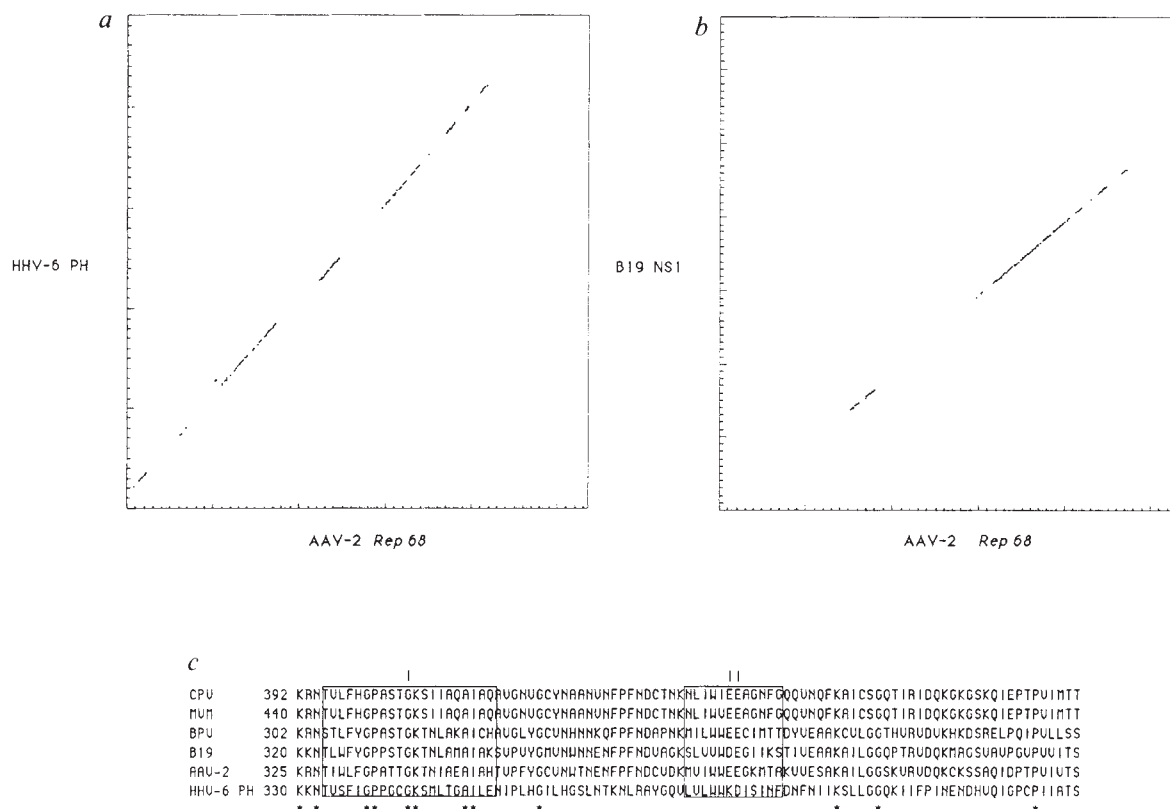


FIG. 3 Comparison of the protein sequences of HHV-6 PH open reading frame and AAV-2 rep 68/78 gene¹⁰ (a) and human autonomous parvovirus B19 NS1 protein²⁸ and AAV-2 rep 68/78 gene (b) using the DIAGON program²¹. c, Multiple alignment of a 95-amino-acid domain in a 145-amino-acid block that is well-conserved in all mammalian autonomous and helper dependent parvoviruses. The sequences used are those of the canine parvovirus (CPV)²⁹, bovine parvovirus (BPV)³⁰, minute virus of mice (MVM)³¹, human parvovirus B19 (ref. 28), the human helper-dependent parvovirus

AAV-2 and the HHV-6 PH open reading frame. The two boxed domains represent motifs common to nucleases, which are separated here by a spacer region of 23 amino acids⁴⁰. Region I contains a consensus purine nucleotide-binding site ((G/A) XXXXGK(S/T))²³. Identical amino acids conserved among all the protein sequences are indicated by an asterisk, and where all but one of the sequences share identical amino acids is indicated by a dash.

repeat¹⁷. The rep 68/78 specifically inhibits DNA amplification³⁵ induced by herpes simplex virus, and cellular transformation by papillomaviruses¹⁸. The C termini of the Rep 68/78 proteins, which are not represented in the HHV-6 homologue, are not required for inhibition of DNA amplification (R. Heilbronn, personal communication). Furthermore, the parvovirus B19 Rep protein, which is less closely related to the AAV-2 rep gene product than is HHV-6 PH, can correctly replicate hybrid

genomes in which AAV-2 origins of replication are substituted for B19 origins³⁶. The conservation between the HHV-6 PH and the AAV-2 rep gene product therefore suggests that HHV-6 PH has a similar range of functions which are advantageous to the survival of HHV-6 in the host. The expression by this ubiquitous herpesvirus of a gene that modulates heterologous gene expression and cellular transformation may have important consequences for the host cell. □

Received 15 November 1990; accepted 15 March 1991.

1. Salahuddin, S. Z. *et al.* *Science* **234**, 596-601 (1986).
2. Briggs, M., Fox, J. & Tedder, R. S. *Lancet* **1**, 40-41 (1988).
3. Okuno, T. *et al.* *J. clin. Microbiol.* **27**, 651-653 (1989).
4. Horvat, R. T., Wood, C. & Balachandran, N. *J. Virol.* **63**, 970-973 (1989).
5. Ensolli, B. *et al.* *EMBO J.* **8**, 3019-3027 (1989).
6. Lusso, P. *et al.* *Nature* **337**, 370-373 (1989).
7. Lusso, P. *et al.* *J. exp. Med.* **167**, 1659-1670 (1988).
8. Takahashi, K. *et al.* *J. Virol.* **63**, 3161-3164 (1989).
9. Gopal, M. R., Thomson, B. J., Fox, J., Tedder, R. & Honess, R. W. *Lancet* **i**, 1598-1599 (1990).
10. Srivastava, A., Lusby, E. & Berns, K. I. *J. Virol.* **45**, 555-564 (1983).
11. Im, D.-S. & Muzyczka, N. *Cell* **61**, 447-457 (1990).
12. Hermonat, P. L., Labow, M. A., Wright, R., Berns, K. I. & Muzyczka, N. *J. Virol.* **51**, 329-339 (1984).
13. Tratschin, J. D., Tal, J. & Carter, B. J. *Molec. cell. Biol.* **6**, 2884-2894 (1986).
14. Labow, M. A., Hermonat, P. L. & Berns, K. I. *J. Virol.* **60**, 250-258 (1986).
15. Labow, M. A., Graf, L. H. & Berns, K. I. *Molec. cell. Biol.* **7**, 1320-1325 (1987).
16. Trempe, J. P. & Carter, B. J. *J. Virol.* **62**, 68-74 (1988).
17. Antoni, B. A. *et al.* *J. Virol.* **65**, 396-404 (1991).
18. Hermonat, P. L. *Virology* **172**, 253-261 (1989).
19. Martin, M. E. D. *et al.* *J. gen. Virol.* **72**, 157-168 (1991).
20. Lawrence, G. L. *et al.* *J. Virol.* **64**, 287-299 (1990).
21. Staden, R. *Nucleic Acids Res.* **14**, 217-231 (1986).
22. Astell, C. R., Mol, C. D. & Anderson, W. F. *J. gen. Virol.* **68**, 885-893 (1987).

23. Li, X. & Rhode, S. L. *J. Virol.* **64**, 4654-4660 (1990).
24. Kozak, M. *Nucleic Acids Res.* **9**, 5233-5252 (1981).
25. Beck, S. & Barrell, B. G. *Nature* **331**, 269-272 (1988).
26. Chee, M. S., Satchwell, S. C., Preddie, E., Weston, K. M. & Barrell, B. G. *Nature* **344**, 774-777 (1990).
27. Moore, K. W. *et al.* *Science* **248**, 1230-1234 (1990).
28. Shade, R. O., Blundell, M. C., Cotmore, S. F., Tattersall, P. & Astell, C. R. *J. Virol.* **58**, 921-936 (1986).
29. Reed, A. P., Jones, E. V. & Miller, T. J. *J. Virol.* **62**, 266-272 (1988).
30. Chen, K. C. *et al.* *J. Virol.* **60**, 1085-1097 (1986).
31. Astell, C. R., Thomson, M., Chow, M. B. & Ward, D. C. *Nucleic Acids Res.* **11**, 999-1018 (1983).
32. Hoggan, M. D., Blacklow, N. R. & Rowe, W. P. *Proc. natn. Acad. Sci. U.S.A.* **55**, 1467-1471 (1966).
33. Cheung, A. K. M., Hoggan, M. D., Hauswirth, W. W. & Berns, K. I. *J. Virol.* **33**, 739-748 (1980).
34. Mendelson, E., Trempe, J. P. & Carter, B. J. *J. Virol.* **62**, 68-74 (1988).
35. Heilbronn, R., Burkle, A., Stephan, S. & Zur Hausen, H. *J. Virol.* **64**, 3012-3018 (1990).
36. Srivastava, C. H., Samulski, R. J., Lu, L., Larsen, S. H. & Srivastava, A. *Proc. natn. Acad. Sci. U.S.A.* **86**, 8078-8082 (1989).
37. Bankier, A. T., Weston, K. M. & Barrell, B. G. *Meth. Enzym.* **155**, 51 (1987).
38. Staden, R. *CABIOS* **6**, 387-395 (1990).
39. Pearson, W. R. & Lipman, D. R. *Proc. natn. Acad. Sci. U.S.A.* **85**, 2444-2448 (1988).
40. Gorbatenya, A. E. & Koonin, E. V. *Nucleic Acids Res.* **17**, 8413-8440 (1989).

ACKNOWLEDGEMENTS. We thank M. E. D. Martin for the HHV-6 clone pH12, J. J. Skehel, M. S. Chee, G. L. Lawrence and U. A. Gompels for advice and discussion, and T. D. K. Brown for critically reading the manuscript. B.J.T. was supported by a Medical Research Council (UK) training fellowship.

Microbiological characterizations by FT-IR spectroscopy

Dieter Naumann, Dieter Helm and Harald Labischinski

Infrared signals of microorganisms are highly specific fingerprint-like patterns that can be used for probing the identity of microorganisms. The simplicity and versatility of Fourier-transform infrared spectroscopy (FT-IR) makes it a versatile technique for rapid differentiation, classification, identification and large-scale screening at the subspecies level.

The detection and characterization of microorganisms by physical techniques promises to be of great value because of the inherent sensitivity, rapidity and potential for complete computerization. Spectroscopic techniques, in particular, are extremely specific and provide a wealth of qualitative and quantitative information about a given sample. The infrared spectrum of any compound is known to express a unique 'fingerprint' and it is this characteristic that enables IR-spectroscopy to be used in the identification of unknown samples using spectral data libraries (for example, in forensic medicine and the pharmaceutical industry). The advent of modern interferometric and Fourier-transform techniques has augmented the specifications of IR-spectroscopy and, thus, paved the way for sophisticated and novel applications of IR in the field of biomedical research.

A joint project was undertaken between 1986 and 1989 to develop efficient methods for the FT-IR analysis of intact bacteria. The project, conducted under the auspices of the Ministry of Research and Technology and the Ministry of Health of Germany, was carried out by researchers at the Robert Koch-Institut of the Federal Health Office in Berlin in collaboration with a manufacturer of FT-IR spectrometers (Bruker, Analytische Meßtechnik, Karlsruhe, Germany). Drawing upon the practical knowledge obtained to date, the prototype of a dedicated instrument, which will be a 'workhorse' for microbiological characterization by FT-IR, is now under development at Bruker and will be available for testing by several laboratories in a joint pilot study in early 1992.

Information basis

The FT-IR technique establishes spectral fingerprints of intact bacteria. These patterns comprise the vibrational characters of all cell constituents, that is, DNA/RNA, protein, membrane and cell-wall components. Consequently, FT-IR probes the total composition of a given organism in a single experiment. As all cell compounds depend on the expression of smaller or larger parts of the genome, the spectra display in a specific way a phenetic and a genetic fingerprint of the microorganisms under study. This is why the selectivity of the new technique is extremely high, allowing differentiations down to the

subspecies and, in most cases, even to the strain and/or serogroup/serotype level¹⁻⁴.

Owing to the multitude of cellular components, broad and superimposed spectral bands are observed within the mid-infrared range (~ 4,000–600 reciprocal wavelengths = wavenumbers [cm⁻¹]). Some spectral sub-ranges are dominated by particular cell components, for example, fatty acids, polysaccharides or proteins. In order to extract specific information from the broad spectral contour, it is advantageous to analyse microbial FT-IR spectra by considering a defined combination of spectral windows and to apply sophisticated mathematical filter techniques (see Fig. 1). As an example of the discriminative power of the method, figure 2 presents the FT-IR classification of some *Pseudomonas* spp. determined by a cluster analysis approach. The dendrogram of figure 2 reveals a hitherto unknown subdivision of the *Pseudomonas fluorescens* sp. in

two main groups, the members of which can be further differentiated down to the strain level within a few minutes of measurement and evaluation.

Experimental procedures

As intact cells are tested, complicated and time-consuming procedures for decomposing cells into single chemical or structural components are avoided. Consequently, simple and uniform procedures are available that are applicable to all bacteria that can be grown in culture; standardization is no longer a crucial step. In short, small amounts of cells (~ 10–60 µg) grown under standardized conditions are transferred from the culture plate to an IR-sample holder using a platinum loop. After a short drying procedure, the sample holder is exposed to the IR-beam, the spectrum is recorded and automatically stored in a data file. The characterization of microorganisms is accomplished

FIG. 1. Whole-cell Fourier-transform infrared (FT-IR) spectra of *Staphylococcus aureus* strain Pelzer. I, Original spectrum, and II, the filtered (resolution enhanced) spectrum as calculated from I. The numbered peaks denote IR bands that can be assigned to specific structural cell components^{1,3,4}.

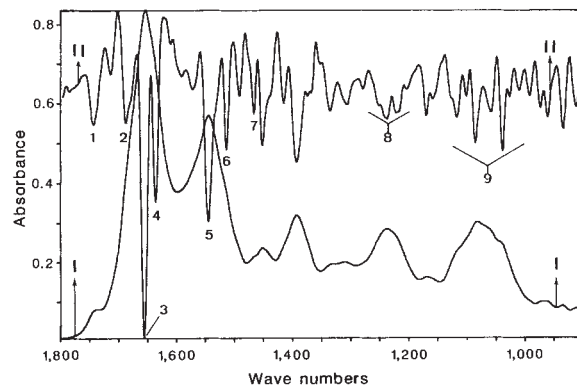
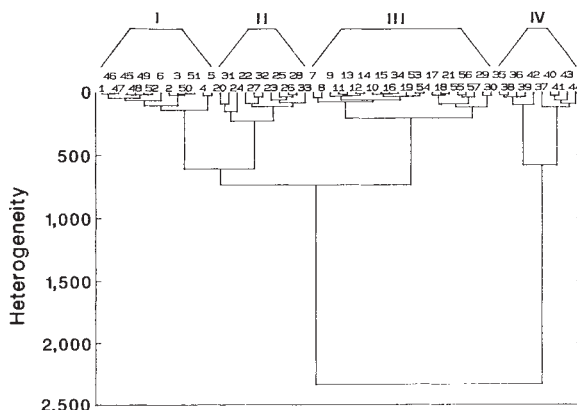


FIG. 2. Dendrogram of the FT-IR classification of a selected series of *Pseudomonas* reference strains (American Type Culture Collection, Rockville, Maryland). I, *Pseudomonas aeruginosa*, II and III, *Pseudomonas fluorescens*, IV, *Pseudomonas cepacia*.



with the aid of comprehensive, computer-stored reference data bases using search algorithms, cluster analysis and multivariate statistics¹⁻⁴. Data evaluation is performed either by a library search, or by a comparison of all spectra collected from identical experimental set-ups. Results are presented by hit lists, dendrograms or by 'factorial maps' displaying graphically the spectral similarities and classifications of the strains under study (see Fig. 2).

The data base of bacterial FT-IR reference patterns that is available in our laboratory comprises spectra from more than 300 very diverse Gram-positive and Gram-negative bacteria. We intend to validate, optimize, and extend the data base and software in the course of the forthcoming joint pilot study. Because of the inherent advantages of FT-IR, it is possible to store data over long periods of time, carry out electronic data transfer to different loci and construct 'home-made' libraries for special user-oriented identifications of unknown species.

Scope of applications

The FT-IR technique is particularly useful for researchers who are engaged in the field of quality and environmental control, as it provides considerable screening power for microorganisms from very diverse human and environmental habitats. It may have application in the recognition of human pathogens in cases where conventional techniques are cumbersome, laborious, inadequate and/or time-consuming, or for which established and routine tests are not yet available. The technology may also help to clarify taxonomic relationships within poorly classified taxa³.

Microbiological FT-IR may also find application in the tracing of epidemics, when used in conjunction with conventional procedures. The capacity for large-scale identity testing may be useful in the maintenance of strain collections or in the supervision of stock cultures to prevent, for example, subtle changes in the production strains used in industry. FT-IR spectroscopy allows the detection of phenotypic consequences of genetic alterations or the analysis of complex, multi-parameter controlled phenomena, such as the production of the bacterial storage material poly- β -hydroxybutyric acid *in situ*³.

The FT-IR approach to sleuthing out the identity of a microorganism was initially developed with bacteria in mind. However, the experimental strategies and the special software packages can easily be extended to the characterization of other microorganisms such as viruses, fungi, yeasts, amoebae, and even mammalian cells.

When applying standard FT-IR techniques, the biomass requirements have been scaled down so that single colonies can be taken from the primary culture plates. Using a special light microscope in conjunction with FT-IR, even microcolony spots of less than 40 μ m in diameter (which corresponds

Microbial matters

A monoclonal antibody to *Helicobacter* spp./*Campylobacter* spp., colony counting systems and a pocket-sized microscope for on-the-spot analysis — product news for the microbiologist.

SAVANT, a specialist in audio-visual teaching aids, has introduced a series of programmes entitled *Protect Yourself from AIDS* (Reader Service No. 101). These audio-visual aids, aimed at laboratory researchers who are at risk from exposure to infected blood and other body fluids, follow safety recommendations laid down by the Centers for Disease Control, the Occupational Safety and Health Administration and the National Committee for Clinical Laboratory Stan-



Audio-visual AIDS.

dards. The video tape, *Protecting Yourself from AIDS: What Everyone Needs to Know* outlines the ways in which HIV is and is not transmitted, together with the causes and symptoms of AIDS. In addition, it reviews the nature and reliability of clinical detection tests for HIV. A second tape, *Protecting Yourself from AIDS: Precautions for Laboratory Workers*, covers safety considerations concerning pipetting, centrifuging and the operation of laboratory equipment, the safe disposal of needles and other sharp instruments, and the clean up and disposal of contaminated wastes.

Reagent round up

Helicobacter spp. and *Campylobacter* spp. can now be easily identified in gastric biopsies using the monoclonal antibody that is now available from Bioproducts for Science (Reader Service No. 102). Data indicate that the monoclonal can be used in the immunocytochemical identification of *Helicobacter pylori* in formalin-fixed gastric biopsies. The monoclonal antibody positively labelled *Helicobacter* spp. in 62.4 per cent of the tis-

to less than 10⁴ bacterial cells) can be analysed. This approach combines, for the first time, the advantages of light microscopy with the sensitivity and selectivity of FT-IR in a single instrument. Microcolony spots that are suitable for FT-IR microscopy are obtained by using an IR-transparent plate as a stamp to transfer the various locally separated microcolonies from the culture plate onto the IR-sample holder. This replica technique permits the detection and enumeration of different colony spots, even from mixed cultures. Optical properties (colony size, colour, different shapes of microorganisms or different colony organizations) are accessible, operator-controlled or by image processing in the computer. As one can easily obtain spectra from 'what you see', the FT-IR pattern technique is used to differentiate and classify individual flora on a primary culture plate.

The software packages used for microbiological characterizations permit the incorporation of additional discriminative, non-spectral characters, such as morphological data from light microscopy or gas chromatography analysis of whole-cell fatty acid composition. The number of spectral traits can be augmented by data that is accessible from the very same FT-IR instrument. The so-called near-infrared and far-infrared regions and, more recently, FT-Raman spectroscopy can

provide a substantial amount of additional and complementary information for dedicated microbiological characterizations.

Any FT-IR instrument can be used for the sensitive and specific monitoring of the various gases (for example, CO₂) that are produced by microorganisms grown in culture. The extra-monitoring capacity of FT-IR can be used to test early bacterial growth in liquid cultures with a sensitivity that is comparable to radioactive isotope tracing⁵.

The increasing use of FT-IR spectroscopy in quality control, microbiological screening, environmental and medical control, and basic microbiological research demonstrates that this technique is a versatile research tool for classifying and identifying microbiological samples. □

Dieter Naumann, Dieter Helm and Harald Labischinski are at the Robert Koch-Institut des Bundesgesundheitsamtes, Nordufer 20, D-1000 Berlin 65, Germany. For more information, fill in reader service number 100.

1. Naumann, D., Fijala, V. & Labischinski, H. *Mikrochim. Acta* **1**, 373-377 (1988).
2. Naumann, D., Fijala, V., Labischinski, H. & Giesbrecht, P. *J. molec. Struct.* **174**, 165-170 (1988).
3. Helm, D., Labischinski, H., Schallehn, G. & Naumann, D. *J. gen. Microbiol.* **137**, 69-79 (1991).
4. Naumann, D., Labischinski, H. & Giesbrecht, P. in *Modern Techniques for Rapid Microbiological Analysis* (ed. Nelson, W.H.) (VCH Verlag Chemie, Weinheim, in the press).
5. Threlkeld, C.H. *J. Food Sci.* **47**, 1222-1225 (1982).

sues from patients with gastritis, says the company. The monoclonal can be used in formalin-fixed, paraffin-embedded tissue as well as cryostat sections. It reacts with the flagellar antigen of most *Helicobacter/Campylobacter* spp.; it does not cross react with other bacteria. The monoclonal is available as lyophilized ascites and is reconstituted to 1.0 ml with sterile distilled water. In addition, a purified FITC-conjugated (fluorescein isothiocyanate) antibody is also available from BPS.

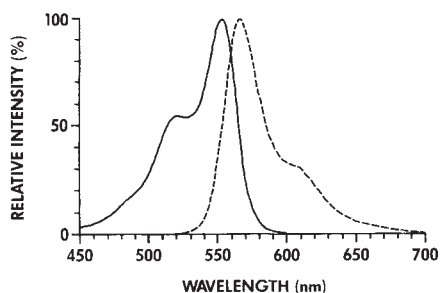
Kirkegaard & Perry has a new **substrate for immunohistochemical staining**, diaminobenzidine (DAB), which produces a sharply localized brown deposit at the reaction site and which is supplied as part of the DAB Reagent Set (Reader Service No. 103). The set includes the DAB substrate concentrate in stable liquid form, buffer concentrate



Diaminobenzidine substrate in kit form.

and hydrogen peroxide solutions. The solutions are provided in dropper bottles that are equipped with controlled-volume tips for ease of use. The need to weigh or handle DAB, a suspected carcinogen, is eliminated as the reagent solution is prepared by dropwise addition, says the company.

Cy3, a cyanine dye developed by Alan S. Waggoner at Carnegie Mellon University and manufactured by Biological Detection Systems, is the first in a new series of fluorophores to be introduced by Jackson ImmunoResearch (Reader Service No. 104). The **fluorophore** is available conjugated to over



Cy3 fluorophore for sale.

270 AffiniPure and AffiniSorbed secondary antibodies, over 60 ChromPure immunoglobulins and their fragments, and streptavidin and avidin. Cy3 is brighter than most other fluorophores, more photostable than fluorescein isothiocyanate, and produces less background than any rhodamine, says Jackson ImmunoResearch.

Frozen/unfrozen assets

Savant Instruments has introduced a **microprocessor-controlled cryopreservation system**, the CPS100, which is designed to increase the viability of stored cell and tissue samples by controlled-rate freezing without the use of liquid nitrogen (Reader Service No. 105). The compact CPS100 unit has a rectangular freezing chamber that measures 15.2 cm × 8.9 cm × 20.3 cm. The two-stage mechanical refrigeration system freezes samples at a selectable rate between 0.1 °C min⁻¹ and 9.9 °C min⁻¹ from 40 °C down to -60 °C, automatically compensating for the heat of fusion as samples change from liquid to solid. Sample temperatures can be monitored continuously by an RTD probe, or thermocouple. Sample racks are available for holding small ampoules, screw-cap vials and straws. The CPS100 can be programmed with up to 20 run protocols.

To complement Pafra's original formula U-Cool 'undercooling' fluid for the sub-zero temperature storage of aqueous solutions, the company has launched U-Cool 'type E' for the **low temperature storage of labile biomolecules** (Reader Service No. 106). Undercooling works as follows: when dispersed in the U-Cool fluid and cooled to -20 °C, aqueous solutions enter a stability window. Even at -25 °C the sample will not freeze, says Pafra. Consequently, undercooling provides all the benefits of low temperature storage, while avoiding harmful freezing. U-Cool 'type E', which contains an emulsifying additive to aid dispersal of aqueous sample in the inert undercooling fluid, is particularly suitable for the storage of enzymes without loss of activity, says Pafra. Costing £65 (UK) per litre, U-Cool and U-Cool 'type E' can be used for the long-term storage of purified biomolecules, purification intermediates, subcellular fractions or microbial cell cultures.

Colony counting

The Mini Light Box from Manostat is intended as a handy **laboratory light source** for screening multiwell plates, Petri dishes,



Manostat's mini light box.

ADVERTISEMENTS

LIPOSOMES

REPRODUCIBLE — HOMOGENEOUS —
EXTREMELY RAPID



Our new LIPOSOMAT guarantees rapid (5ml within 60 minutes) preparation of uniformly sized unilamellar liposomes. Liposome size is selectable between 25 and ca. 600 nm diameter. Sample volume is normally 5-10ml or up to 200ml if used in combination with an additional instrument. Lipid concentration can be up to 300 mg/ml.

DIANORM

POB 126, D-8000 Munich 65, FRG

Reader Service No.17

TIB MOLBIOL BERLIN
CUSTOM DNA SYNTHESIS

DM per base
5,-

DM 95,- per oligo

FAX (0049) 30 6261960

Reader Service No.79

Synthetic Oligonucleotides

Custom made probes and primers,
synthesised, purified and shipped in 3 days

£3.50 per base

(All inclusive)

0.2 µM scale (approx. 500 µg)

Modified bases, biotin or digoxigenin labelling,
data base analysis and full design facilities

Molecular Medicine Unit,
King's College School of Medicine & Dentistry,
Denmark Hill, London SE5 8RX

Tel. XX-44-(0)71 - 326 3126

FAX XX-44-(0)71 - 733 3877

Reader Service No.56

Professional Custom Services

DNA SYNTHESIS
PROTEIN SEQUENCING
PEPTIDE SYNTHESIS

Rapid Service • Guaranteed Quality • Ample Quantity
Competitive Prices • Free Quotations
World-wide Delivery

V.G. Services 1200 Wonderland Rd. S. Bldg #7, Unit #2
London, Ontario Canada N6L 1A8
CALL/FAX: (519) 652-3758 ALT. FAX: (519) 652-9622

Reader Service No.16

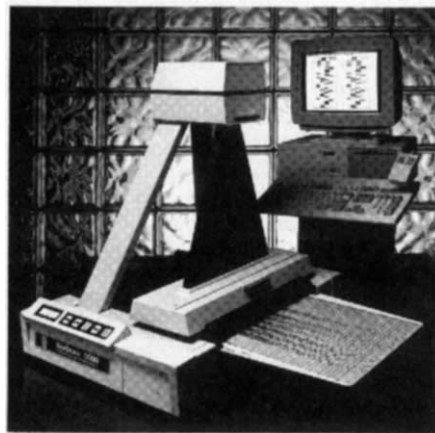
electrophoresis gels and slides (*Reader Service No. 107*). The \$79 (US) battery-operated light box features a colour-corrected fluorescent light that is evenly diffused over the 4 × 5-inch viewing area. A 6 V d.c. adaptor socket is provided for stationary use of the light box. An optional \$79 (US) 1.7× magnification Mini Magnifier is available to assist in the counting of bacterial colonies and hybridoma screening. Removable counting grids, which cost \$10 (US) for a pack of five, provide a useful reference when counting colonies. Manostat offers a complete colony counting system, which consists of the Mini Light Box, Mini Magnifier, and colony counting pen, for \$300 (US).

Domino, the new colony counting system from Perceptive Instruments, is designed to provide accurate plate counts in seconds and has comprehensive data handling facilities (*Reader Service No. 108*). The system, which can be fully integrated with IBM AT-compatible computers, comprises the image analysis hardware, a high-performance video camera and the Borland Quattro Pro 2 spreadsheet program. Other design features include a modular camera stand, which provides a choice of illumination for counting colonies on different culture media, and a zoom lens attachment for laboratories using different sized Petri dishes or those requiring higher magnifications. Other features include special algorithms to separate touching colonies and mouse-controlled frames to isolate spreaders. The count mode is useful for rough-edged colonies that are often

found in settle plates or mammalian cell culture. Domino can also differentiate colonies on the basis of size. Other applications of Domino include antibiotic assays, epi-fluorescence microscopy, multipoint sensitivity testing and toxicological assays.

Mastscan Series 600 is a new **automated reading and reporting system** from Mast Diagnostics for bacterial antibiotic susceptibility and identification tests (*Reader Service No. 109*). The system permits the microbiologist to analyse multipoint inoculated susceptibility and identification plates with great speed and accuracy, says Mast. Different applications — breakpoints, automatic identifications, M.I.C. determinations and epidemiological analysis — are selected by means of menu displays. The hardware consists of a multipoint inoculator, which is designed for the delivery of 19, 36 or 96 standardized bacterial inocula in a rapid and precise manner, a scanner module (high-resolution CCD colour TV camera), a microcomputer and colour monitor.

Automate DNA sequence reading and other laboratory tasks with the SciScan 5000 **automated scanning system** from United States Biochemical (*Reader Service No. 110*). The high-resolution scanner digitally records an image of any laboratory sample, including wet gels and culture dishes, with a scanning bed that accommodates samples of any length, up to 14 inches wide and 0.5 inches thick. USB's BioAnalysis software controls the scanner, displays the scanned images on the computer screen in 'windows', and analyses the images. Analysis can be simple (densitometry and graphing) or complex (DNA sequence reading). The software



USB's automated scanning system scans films, wet gels and culture dishes.

can perform whole-band densitometry with automatic or manual lane and band finding. Stored, digital images of laboratory samples can be modified to enhance contrast, remove background, soften, sharpen or even 'shadow' an image. Each of these functions allow the user to view important aspects in an image while minimizing background. Images can be printed or saved in a standard file format for processing by other programs.

Under the microscope

A new automated system for **microinjection into living cells** has been developed by Carl Zeiss (*Reader Service No. 111*). The AIS (automated microinjection system), which is based on the new Zeiss Axiovert inverted microscopes, is designed to provide computer-controlled injections into as many as 2,000 cells per hour. The long working distance ICS optics and condensers of these microscopes provide ample space above the specimen for injection tools, says Zeiss. Control over the injection procedure is provided



Computer-controlled microinjection.

by the Microscope System Processor, which features both hard and floppy disk drives and VDU graphic controller. Zeiss says the program controls the positioning of the scanning stage in both the X and Y axis with sub-micron reproducibility, together with the motorized movement of the micromanipulator in the Z axis, and the CCD camera. The position of single cells can be selected and marked using a menu-selected graphic overlay program, which stores the relevant XY coordinates for rapid retrieval.

The Lensman **pocket-sized microscope**, which is distributed by Inlet, measures just 105 mm (diameter) × 27 mm (thickness) and can be used to perform microscopic examinations anyplace, anytime (*Reader Service No. 112*). The top surface forms the specimen stage, while the operating controls are placed on the underside. The nine-element folded optical system provides magnifications of 80× and 200× and a resolving power of 4 μm. Slides or opaque objects are placed on top of the Lensman, allowing the light source, which can be set at high or low power, to illuminate the specimen. Part of the top case articulates to form a lighting arm with a range of illuminations: reflected, dark-ground and transmission. As an optional extra, the Lensman can be fitted with a Photo-adaptor, which allows direct coupling to a 35-mm SLR camera. □

These notes are compiled by Diane Gershon from information provided by the manufacturers. To obtain further details, use the reader service card bound inside the journal. Prices quoted are sometimes nominal, and apply only within the country indicated.

ADVERTISEMENT

DNA

Around the World

\$4.50

a base
(no set-up)

Includes:

- 500-1500 ugs
- 15-45 O.D.S.
- 24-72 hour delivery
- Amplification ready
- Gel picture
- Quality guaranteed products

We are a full service and design DNA lab.

1-800-227-0627

Fax 1-214-420-0442

BIO-SYNTHESIS

Reader Service No.18